LANGUAGE BIAS IN VISUALLY DRIVEN DECISIONS

computational and neurophysiological mechanisms

Lara Todorova

The educational component of the doctoral training was provided by the International Max Planck Research School (IMPRS) for Language Sciences. The graduate school is a joint initiative between the Max Planck Institute for Psycholinguistics and two partner institutes at Radboud University – the Centre for Language Studies, and the Donders Institute for Brain, Cognition and Behaviour. The IMPRS curriculum, which is funded by the Max Planck Society for the Advancement of Science, ensures that each member receives interdisciplinary training in the language sciences and develops a well-rounded skill set in preparation for fulfilling careers in academia and beyond.
More information can be found at www.mpi.nl/imprs.

The research in this thesis was supported by Gravitation Grant 024.001.006 of the Language in Interaction Consortium from Netherlands Organization for Scientific Research.

LANGUAGE in INTERACTION

NWO
Netherlands Organisation for Scientific Research

ISBN
978-94-6284-243-4

Cover design
Lara Todorova and Promotie in Zicht

Design/layout
Promotie In Zicht (www.promotie-inzicht.nl)

Print
Ipskamp Printing, Enschede

LANGUAGE BIAS IN VISUALLY DRIVEN DECISIONS

computational and neurophysiological mechanisms

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college van decanen
in het openbaar te verdedigen op dinsdag 9 februari 2021
om 16.30 uur precies

door

Lara Todorova
geboren op 11 januari 1988
te Moskou, Rusland

Promotor
Prof. dr. P. Hagoort

Copromotoren
Dr. V. Piai
Dr. D.A. Neville

Manuscriptcommissie

Prof. dr. R. Cools	
Prof. dr. B.U. Forstmann	*Universiteit van Amsterdam*
Dr. J.C. Francken	*Universiteit van Amsterdam*

Table of Contents

1

Introduction

It takes less than half a second to recognize your PI as the person next to the coffee machine. Usually it is a fast and effortless task. It starts with an image being projected on the retina of the eyes, which in its turn starts the recruitment of a cascade of visual regions and ends with us reaching a decision about whom or what we see (Figure 1 A). Visual perception, however, is not only influenced by the features (dark hair, brown eyes, etc.) that we process and combine together to form a unified percept (PI, for example). Prior information or beliefs one might have are known to influence visual perception too, and this influence can come in different flavors: experience (Figure 1, B), context (Figure 1, C) or cues (Figure 1, D). In the coffee machine example, prior information can lead to wrong recognition, for example, if you know the PI often hangs out at the coffee machine. If you see someone with those features, you tend to expect it is most likely your PI (which could be wrong because there might be another person with similar features who is also often at the coffee machine).

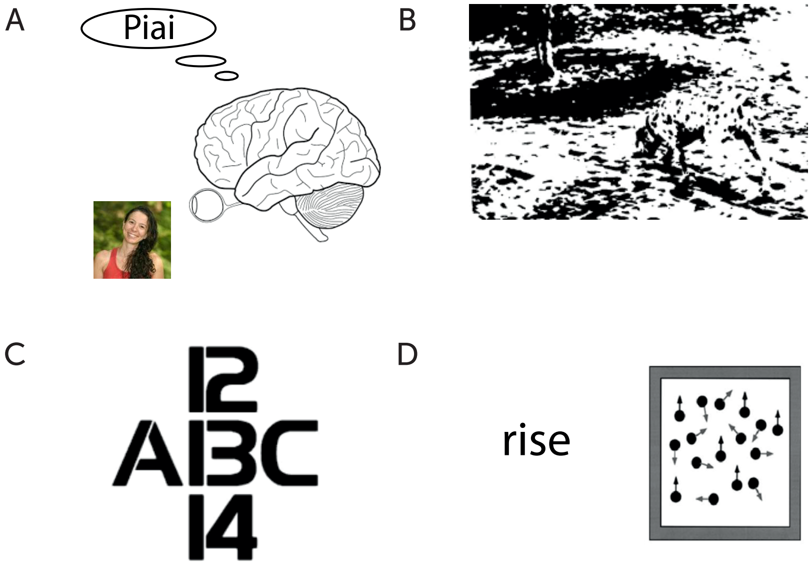


Figure 1. Bottom-up (A) An image being projected on the retina of the eyes starts recruiting a cascade of visual regions and ends with us reaching a decision about whom or what we see. **Top-down (B-D) visual perception**. Influences of prior information on perception: **B)** Experience effects: If you look at the picture, it is far easier to notice a dalmatian there if you have lots of experience with dogs. **C)** Context effects: depending if you pay attention to the letters (A, C) or the numbers (12, 14) you will perceive the middle symbols as a letter (B) or a number (13). **D)** Cue effects: You are faster and more accurate to decide about the direction of the dot motion if the word describing the direction of the moving dots ("rise") corresponds with the direction of the moving dots (most of the dots moving up-wards).

For humans, words are important cues that we use for many reasons, and making things more clear is one of them. A simple instruction such as "now you will see a picture of a dalmatian" will help you to decide what you are about to see is (most likely) a dalmatian. Even if the things we see are not ambiguous or hard to perceive, words help us notice these things on time. For example, when seeing the road sign "Slow traffic ahead"/ "Fresh oil" (Figure 2) one is more likely to pay attention to the road, so she or he can make better decisions.

Figure 2. Road signs as an example of how words can affect our behavior.

What does it mean, we make better decisions? On the road it means we are faster to notice potentially dangerous things (slippery road) and react appropriately (not make any sudden moves). In other words, we process information faster and more efficiently. How can we get insight in how information is processed by looking only at reaction times and choice responses? And how does our brain compute the visual decisions affected by words? These are some of the questions that I address in this thesis.

Sometimes words can serve not as attention pointers, like in the situation with the traffic signs, but as attention templates where we have to match the meaning of the word to the meaning of an image. Imagine, for example, that you are following a recipe and you have to pick a witlof, ignoring, for example, an avocado. It is a common finding that targets with semantically related words (we read "witlof", see an avocado, and have to pass on the avocado, because it is obviously not a witlof) would take us more time to process in comparison to the targets with semantically unrelated words (we search for a spoon, see an avocado, and have to pass on the avocado because it is not a spoon). To succeed in this task, we have to resolve semantically similar representations. In this thesis I also examine a potential computational mechanism of an attention-template type of effect, relating it to the *controlled semantic cognition* framework (see further below).

First, I will introduce visual perceptual decision making as a concept, capitalizing on one of the most influential computational frameworks of decision making – the drift diffusion model (DDM). Afterwards, I will illustrate how this approach can shed light on the mechanism of language based decision making, using behavioral and neuroimaging experiments.

Perceptual decision making

Perceptual decision making refers to the process of choosing an option based on available sensory evidence (see Figure 3 A). Visual categorization can be considered as a specific case of perceptual decision making: we refer to it as the ability to assign sensory stimuli to discrete groups that are behaviorally relevant (Freedman & Assad, 2016). For an example of visual categorization see Figure 3 B.

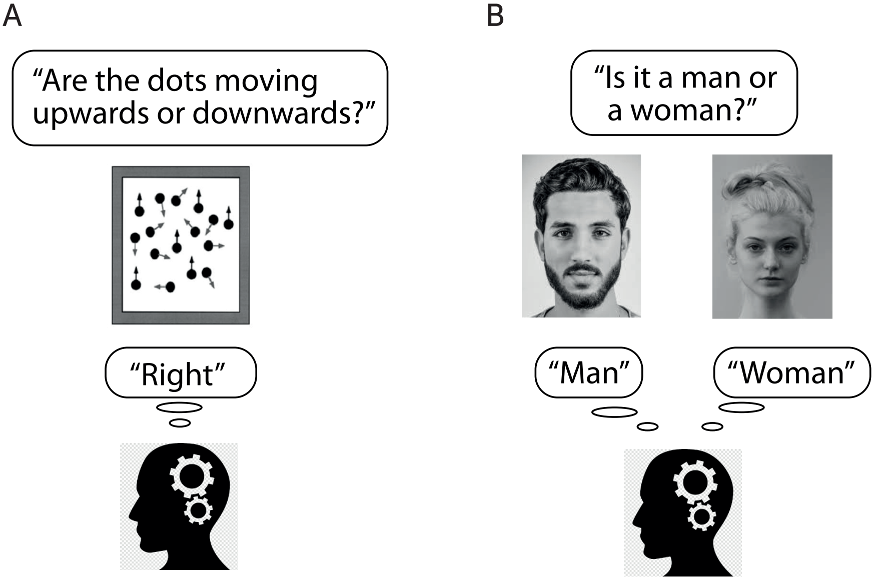


Figure 3. A) While presented with the display where the dots move to the left or right, the person decides about the direction of the movement (left or right). Depending on the coherency of the movement, the decision about it can be easy (for example if the majority of the dots are moving right, it is easier to decide that they move to the right) or difficult (for example if it is not obvious that he majority of the dots is moving right, it is more difficult to decide that the dots move to the right). **B)** When one perceives a woman or a man when looking at the person's face, she or he performs a binary decision, i.e., she or he decides whether the person is a man or a woman based on the visual features.

Regardless of what we decide about (low-level percepts such as the direction of the dot motion or more high-level perception such as the category of an animal), the process of perceptual decision making can be described as a sequence of processing stages (sensory, decision formation and motor execution) that unfolds over time (Figure 4).

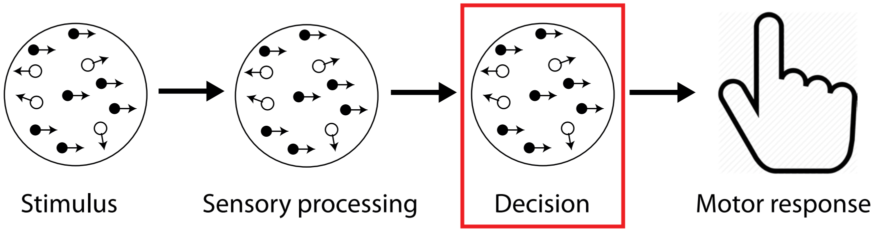


Figure 4. Decision making as a process. When we are presented with a cloud of randomly moving dots, first, we encode sensory information, then make up our mind about this information, and finally respond.

Well-known theories of decision making, such as Signal Detection Theory or Probabilistic Population Codes [BOX1], are focused on describing the *decision formation phase* (red square in Figure 4) in the decision making process. Therefore the contributions of the sensory and motor systems are disregarded.

[BOX 1] Signal Detection Theory or Probabilistic Population Codes

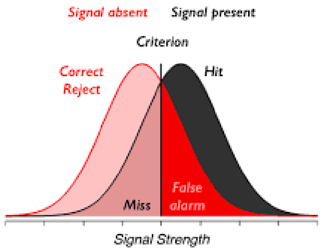
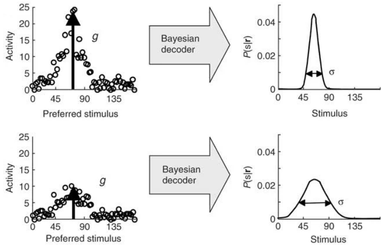
Signal detection theory (SDT) describes at which point we are able to detect a signal in the environment in terms of stimulus intensity and decision threshold placement. Let's consider an example where we describe a process of recognizing your PI via a signal detection framework (i.e., PI is a "signal"). When seeing a person next to the coffee machine with her back to you, it is not clear whether she is your PI. There might be different scenarios: it can indeed be your PI and you correctly greet her: «Hi, Vitoria![1]» (that is a hit!). It can be that it was Vitoria but you are deeply focused on your own thinking and say nothing (that is a miss!). It can be that it is not Vitoria, and you say «Hi, Vitoria!» (it is a false alarm!). It can be that it is not Vitoria and you asocially say nothing (that's a correct rejection).

---

[1] Vitoria Piai is the PI in the "Language Function and Dysfunction" lab at the Donders Institute. She never hangs out at the coffee machine, so it is unlikely to develop this prior about her.

Each choice (saying hi/not saying hi) gives rise to a normal distribution over many trials, i.e., multiple meetings at the coffee machine, and I give a response based on a strategy and strength of visual information (i.e., evidence). For example, I can say «Hi Vitoria» to all people remotely resembling Vitoria (liberal strategy), or I could greet her only when I am very sure it is her (conservative strategy). Decision strategies are defined in terms of a decision threshold, i.e., criterion. Setting it high would mean I adopt a conservative behavior, conversely a low threshold would mean I adopt a more liberal behavior. Regardless of the strategy, it can be just easy to notice Vitoria due to some circumstances like weather conditions or good light in the hallway (good quality of visual information results in larger perceptual sensitivity) as opposed to situations when it is hard to notice her (poor quality of visual information results in lower perceptual sensitivity index). The process of making a decision is therefore a process of estimating the intensity of the stimulus and placing the criterion at a certain place in the distribution. To sum up, Signal Detection Theory (see Table 1) describes decision making in terms of behavioral strategies that are useful for performance. STD however is not a model of cognition, which means that its parameters do not reflect aspects of cognition.

Table 1 Brief illustration of Signal Detection Theory vs. Probabilistic Population Codes approach to representing decisions..

Signal Detection Theory		2-choice decisions	-bias (c) - perceptual sensitivity (d’)
Probabilistic Population Code		Continuous Quantity decisions	Full distribution

The picture illustrating probabilistic population code is taken from (Ma, Beck, Latham, & Pouget, 2006)

Probabilistic population codes (see Table 1) assume that a population of neurons produces a neural response that represents the probability of occurrence of a certain event (e.g., orientation of a grating, speed of a moving dot or the odds of seeing your PI). From this point of view, our choices are probabilistic decisions that are dependent on our previous experiences and current context. The process of making a decision is therefore an act of probabilistic inference. I keep this information here brief in order to highlight the differences between probabilistic and sequential inference.

The work on this thesis used sequential stochastic evidence accumulation as a computational framework.

In contrast, the drift diffusion model (DDM) [BOX 2] formalizes sensory processing and motor response (described as non-decision component) and factors them out from the process of the *decision formation phase.* While it is true that in the DDM it is not possible to disentangle motor vs. sensory contributions, the advantage of this model is the ability to disentangle decision vs. non-decision components, which makes the process of *decision formation* less contaminated by the processes of motor execution and stimulus encoding. DDM is a well-known tool in the field of decision making when it comes to quantifying force-choice binary tasks. In the field of language, this is quite an avant-garde approach, the pros and cons of which I discuss in behavioral (Chapter 2), neuroimaging (Chapter 3, 4) and neuropsychological (Chapter 5) studies.

[Box 2] Drift diffusion model

The mystery and challenge of the human mind is that its cognitive processes cannot be directly observed. We have to take indirect measures of latent processes such as attentional focus or memory formation. Traditionally, human behavior is measured via reaction time (RT) and percentage of correct responses, which were measures originally developed for studying the speed of mental processes (Franciscus Donders) and the effects of physical magnitudes of stimuli along with the sensations they produce (Gustav Fechner). Using only simple behavioral measures (RTs and correct responses) we cannot make inferences about specific cognitive processes that underlie human behavior. To address that, one needs to consider that certain behaviors can be the result of several cognitive processes and nuisance processes. In other words, the task has to be represented by characterizing every relevant cognitive process, which in turn requires the use of cognitive process models (Forstmann & Wagenmakers, 2015).

Models of cognitive processes quantify psychological processes in terms of parametric descriptors. Behavioral data (RTs and responses) is used to estimate the parameters of the model, which allow inferences on the hidden (i.e., latent) psychological processes that underlie the task. The work presented in this thesis relies heavily on cognitive modeling, specifically on one particular family of models – the Drift Diffusion Model (DDM, for review see Ratcliff & McKoon, 2008; Smith & Ratcliff, 2004; Usher & McClelland, 2001). In DDM, the process of making a decision is described as a stochastic accumulation of evidence towards a response. The parameters of the DDM model that characterize this process are: decision threshold, drift rate, starting point and non-decision time (see Figure 5). Each of them bears cognitive interpretability, which has been shown in both human and animal research (for review see Mulder, van Maanen, & Forstmann, 2014; Ratcliff & McKoon, 2008). A simple process of deciding whether the dots are moving to the left or right can be described in terms of the factors that underlie the performance: (1) the features of the stimuli have to be encoded (the perceptual machinery can take varying time to do so); (2) the perceptual evidence has to be accumulated efficiently enough; (3) the decision threshold level has to be set appropriately for a task; (4) the motor response is executed after a decision has been made. Here, I clarify certain assumptions of the model. First, the model assumes a certain sequence of decision making phases: encoding (I), evidence accumulation (II), motor execution (III). Second, the model does not differentiate the nature of the non-decisional component: encoding (I) and motor execution stages (III) are agnostic to the model, which is reasonably reflected in the name of the parameter "non-decisional" component.

Hierarchical drift diffusion model

In this thesis I use a hierarchical variant of Drift diffusion model, HDDM (Wiecki, Sofer, & Frank, 2013). The hierarchical model allows estimating parameters at both individual and group level by assuming that participants within each group are similar but not identical to each other. This approach is conceptually different from the traditional approach, where psychological models are fit separately to each individual (assuming that subjects are completely independent of each other) or to the group (assuming participants are all the same). Traditional approaches are suboptimal because they fail to either capture differences between the subjects or miss out on statistical strength. In HDDM subjects are considered to be drawn from the group distribution, which flexibly constrains the individual parameters within the group (see Figure 6). This results in HDDM

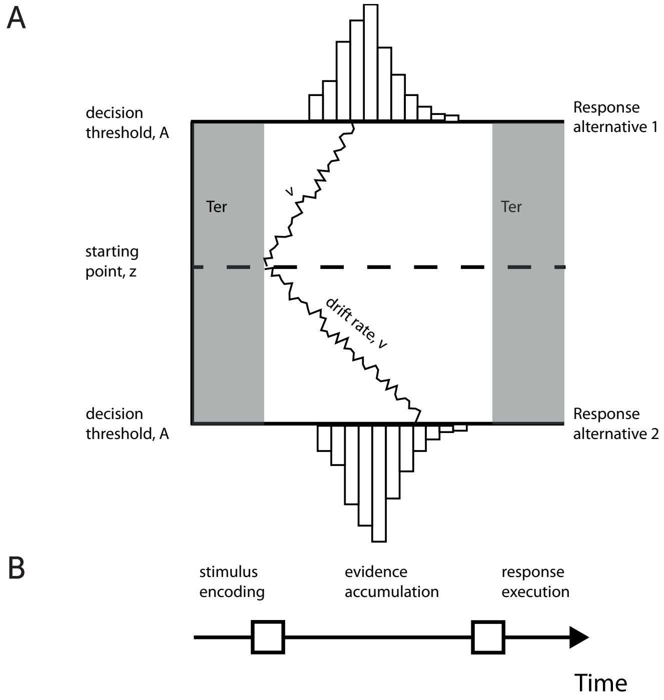


Figure 5: (A) The Drift Diffusion model with the four parameters: drift rate (*v*), decision threshold (*A*), starting point (*z*), non-decision component (*Ter*). **(B)** Stages of processing characterized by the DDM model. Different stages of processing are highlighted by the shading in the panel A to highlight the mapping of the DDM components to the processing stages.

requiring less data per subject and condition to estimate the model in comparison to non-hierarchical methods. This is an essential feature for studies where collecting hundreds of trials is not feasible, both in clinical settings (such as Chapter 5, where patients are involved), and in the lab settings where experimental time is limited.

Graphical nodes are distributed as follows: μ_a ~ **G** (1.5, 0.75), σ_a ~ **HN** (0.1), μ_v ~ **N** (2, 3), σ_v ~ **HN** (2), μ_z ~ **N** (0.5, 0.5), σ_z ~ **HN** (0.05), μ_t ~ **G** (0.4, 0.2), σ_t ~ **HN** (1), a_p ~ **G** (μa , $\sigma 2a$), v_p ~ **N** (μv , $\sigma 2v$), z_p ~ invlogit (**N** (μz , $\sigma 2z$)), t_p ~ **N** (μt , $\sigma 2t$), and $x_{p,\ j}$ ~ F(a_p , z_p , v_p , t_p) where $x_{p,\ j}$ represents the observed data, which consists of response time and choice of subject p on trial j and F represents the DDM likelihood function as formulated by (Navarro & Fuss, 2009). **N** represents a normal distribution parameterized by mean and standard deviation, **HN** represents a positive-only, half-normal parameterized distribution by standard-deviation, **G** represents a Gamma distribution parameterized by mean and rate, **B** represents a Beta distribution parameterized by α and β. HDDM then uses Markov chain Monte Carlo (MCMC) (Gamerman and Lopes, 2006) to estimate the joint posterior distribution of all model parameters. Crucially, the joint posterior distribution is conditioned upon a specific model, therefore it is important to touch upon principles of model construction and model selection. For the model construction, we varied A, Ter, v (set to fixed or free) over the experimental factor, which defined the number of possible models. For example, in Chapter 4 we had one experimental factor, which with all possible combinations of free/fixed parameters (A, Ter, v) resulted in 7 models. In Chapter 3 there were two factors, which were set as free or fixed across three parameters (A, Ter, v) and resulted in 19 models.

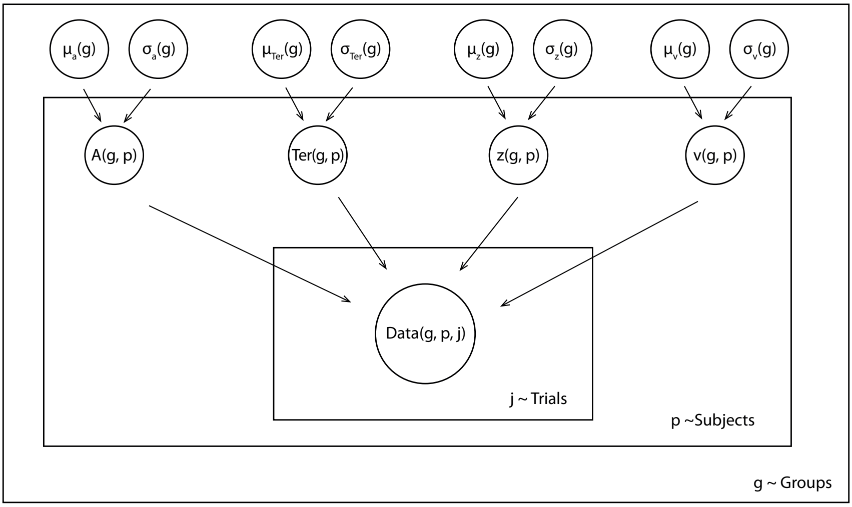


Figure 6: The graphical representation of the hierarchical model estimation. The node Data (*g,p,j*) indicates the observed data of each participant group (*g*), participant (*p*), and trial (*j*). Nodes *a* , *T er* , *z, and v* are parameters of the drift-diffusion model (at the single subject level), each with a group distribution for each participant group with mean *μ* and standard deviation *σ*.

For the model selection, we relied on the deviance information criterion (DIC), which is a common method for assessing model fit in hierarchical models (Spiegelhalter, Best, Carlin, & Van Der Linde, 2002). The model with the smallest DIC is estimated to be the model that would best predict the currently observed experimental dataset. Therefore we would consider a lower DIC as an indication towards better fit of the model, and the model with the lower DIC would be considered a model-winner in the given model set. Furthermore, alternative measures of model selection such as Bayesian Information Criterion (BIC) or Akaike Information Criterion (AIC) are metrics that rely on maximum likelihood estimation and therefore are not appropriate in the case of MCMC fitting routines that rely on Bayesian statistics. Thus, DIC is an appropriate metric for the present purpose to compare models in their ability to capture the observed data with the minimal amount of complexity required. Importantly, all models in this thesis were inspected for convergence both visually and quantitatively. For the visual part, I examined that the posterior samples ("the chain") of a given model have properly converged. I examined whether a) there are no drifts or large jumps in the trace; b) the autocorrelation (i.e. the influence of past samples) drops to zero rather quickly (i.e. smaller than 50). For the numerical part, I assured that a MCMC error for all of the parameters is smaller than 0.01 and an R-hat statistic (which is a measure of convergence among multiple MCMC chains) is under 1.1. The model with the lowest DIC that passed all the convergence checks was considered the model-winner. Further, for the best fitting model, I examined the differences in posterior estimates by conducting statistical analyses. To highlight, posterior estimates of model parameters are of ultimate interest for interpretation of the results in this thesis.

According to one of the most prominent theories of decision making, in order to reach a perceptual decision, the brain has to accumulate noisy sensory information over time until the decision threshold is reached (Gold & Shadlen, 2007; Mazurek, Roitman, Ditterich, & Shadlen, 2003; Ratcliff, Smith, Brown, & McKoon, 2016). One of the most important findings illustrating evidence accumulation is the persistent neuronal firing in the lateral intraparietal area (LIP) during a motion discrimination task. This neuronal firing persists until a critical threshold is reached and the decision process is terminated, followed by the response execution (Kim & Shadlen, 1999). LIP is not the only brain area that can accumulate evidence for decision choice. In auditory (Binder, Liebenthal, Possing, Medler, & Ward, 2004), somatosensory (Pleger et al., 2006) and visual (Heekeren, Marrett, Bandettini, & Ungerleider, 2004) decision making tasks, candidate brain regions for evidence accumulation were found in anterior cingulate and dorsolateral prefrontal cortices using correlations between the blood oxygenation level-dependent (BOLD) signal and trial-to-trial reaction times (RTs). To illustrate, various brain regions (Figure 7) can be functionally involved when making a decision. Importantly, these regions reflect the mechanisms implemented in the stochastic models of perceptual decision making.

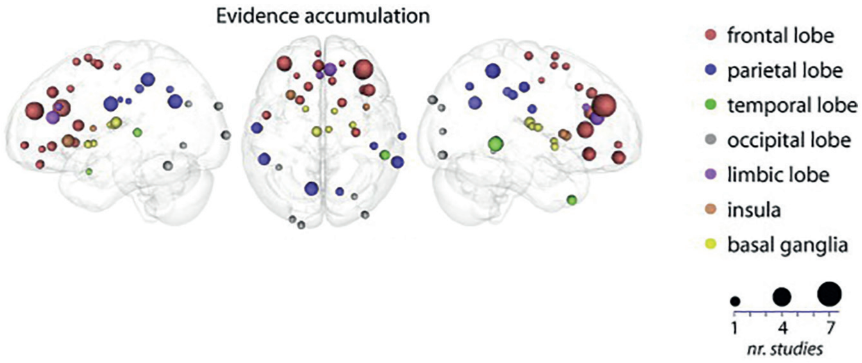


Figure 7: Brain map renderings of the peak coordinates reported by the studies that include evidence accumulation (drift rate) parameter in the fMRI analysis (reprinted and modified from Mulder, van Maanen, & Forstmann, 2014).

Words as top-down cues for visual decisions

As I mentioned in the beginning of the chapter, the brain actively constructs what we see based on information prior to perception (top-down information), such as expectation or attention (Gilbert & Sigman, 2007; Rauss & Pourtois, 2013). For example, verbs congruently describing the direction of upcoming random-dot motion kinematograms (e.g., word 'rise' followed by the dots moving up-wards) result in faster and more accurate judgments about the presence or absence of the motion itself (Meteyard, Bahrami, & Vigliocco, 2007). In spite of the fact that the effect of language on perceptual decision making is well documented (Boroditsky, 2001; Winawer et al., 2007, see for review Raftopoulos & Lupyan, 2018), the locus of this effect is still a matter of investigation. While some studies showed that language can modulate perceptual representations (Amado et al., 2018; Puri, Wojciulik, & Ranganath, 2009), others suggest that language rather involves processes of lexico-semantic memory and/or decision making (Francken, Kok, Hagoort, & de Lange, 2015; Tan et al., 2008) via which it influences visual decisions. In short, it remains unclear how and at which processing stage language cues (i.e., words briefly presented before the visual input) affect perceptual decision making. To advance our knowledge regarding the functional role of brain regions reflecting the effect of language on perceptual decision making, I tested whether brain activity in those brain areas can be explained in terms of evidence accumulation.

It is important to emphasize, that the goal of this model-based approach is not finding a brain region that would reflect a model parameter (Mulder et al., 2014). The goal is rather decomposing behavioral performance into processes and testing whether brain activity can reflect this process. In this sense, the model parameters that represent cognitive states do not have to necessarily map onto a neural substrate. Instead, they provide an intermediate level between behavior and neural levels that can be meaningful for interpreting the neural activation functionally. DDM can be useful in contributing to already existing theoretical frameworks (such as semantic control, see further below). Whereas much research has been done on decision making on the one hand and semantic cognition on the other, they have not often been put together. In this thesis, I fill this gap by bringing new insights to the field of semantic cognition and language from the computational perspective point of view.

Semantic control from the decision making performance perspective

Broadly speaking, semantic cognition refers to the ability to "use, manipulate and generalize knowledge". Semantic cognition supports a broad range of verbal (e.g., naming pictures or selecting a word among synonyms to describe a picture more accurately) and non-verbal (e.g., using tools in creative ways) behaviors (Lambon Ralph, Jefferies, Patterson, & Rogers, 2017). It has been proposed that semantic cognition is subserved by two highly interactive neural systems: the systems of *semantic representation* and *control*. The system of *semantic representation* relies on the hub-and-spokes neural architecture with hubs localized in the anterior temporal lobes (ATL) and spokes placed in modality associated cortices (see **Figure 8, A**).

The system of *semantic control (***Figure 8, B**) consists of a system of brain regions that are separate from the *semantic representation* system. The system of semantic control interacts with the semantic system by boosting some aspects of semantic knowledge when it is weakly encoded (for example, in tasks that require focus on non-typical features). Most computational theories of language production and comprehension accept that in addition to representational (i.e. lexico-semantic) processes, executive mechanisms govern the extraction of relevant semantic information from stimuli (Badre, Poldrack, Paré-Blagoev, Insler, & Wagner, 2005; Jefferies & Lambon Ralph, 2006; Lambon Ralph, Jefferies, Patterson, & Rogers, 2017; Nagel, Schumacher, Goebel, & D'Esposito, 2008). Broadly speaking, executive control comprises of both semantic working memory retrieval and selection between competing representations (Engle & Kane, 2003). The exact mechanisms

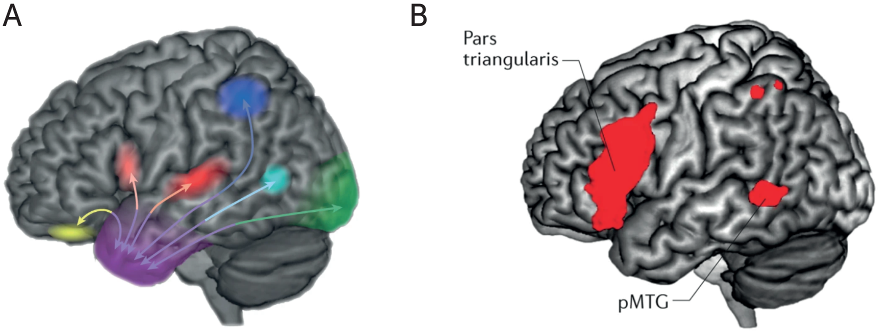


Figure 8: Systems of semantic representation (A) and control (B) (adopted from (Lambon Ralph et al., 2017).

of executive control in the semantic domain are still a matter of investigation. Furthermore, it is still debated to what extent the system of semantic control overlaps with executive mechanisms. Therefore, a computational account of semantic cognition that would account for executive/control functions would be highly informative.

Previous studies suggested that evidence accumulation may be considered as a descriptive model of participants behavior that could inspire theories of language retrieval, and indirectly language control (Anders, Ries, van Maanen, & Alario, 2015; van Maanen & van Rijn, 2007). For example, it has been shown that patients with lesions in left prefrontal cortex (lPFC) failed to adjust the decision threshold for lexical selection in language production (Anders, Riès, Van Maanen, & Alario, 2017). This finding not only contributed to the idea that semantic representation and control are dissociable systems, but also proposed a viable mechanism for semantic control. In the field of decision making it has been shown that the decision threshold modulation is essential for exhibiting the necessary amount of control to solve the conflict efficiently (Cavanagh et al., 2011; Forstmann et al., 2010). These results suggest that evidence accumulation is a fruitful framework to reveal various aspects of executive control mechanisms. In Chapter V I investigated the mechanisms of semantic representation and control via the lens of DDM.

Thesis outline

This dissertation includes four empirical chapters that addressed different questions regarding the mechanisms of visuo-semantic interactions. I used a combination of behavioral, computational, neuroimaging and neuropsychological methods to address this topic. Particularly, I formalized visuo-semantic interactions from a decision making perspective, i.e., as a process of evidence accumulation to the decision threshold. In Chapter II I show, as a proof-of-concept, that the language advantage in visually driven decision making translates into the speed of evidence accumulation. In Chapters III and IV, I investigated whether neural markers underlying linguistically boosted visual decisions can be accounted for by the speed of evidence accumulation or by any of the others DDM parameters (i.e., decision threshold, starting point). Chapter V proposes an evidence accumulation framework as an exploratory platform for disentangling the processes of semantic control and interference in the word-picture matching task.

2

Associative and identity words promote the speed of visual categorization: a hierarchical drift diffusion account

A modified version of this chapter has appeared as: Todorova L. & Neville D.A., (2020). Associative and identity words promote the speed of visual categorization: a hierarchical drift diffusion account. *Frontiers in Psychology*, 11, 1-17.

Abstract

Words can either boost or hinder the processing of visual information, which can lead to facilitation or interference of behavioural response. We investigated the stage (response execution or target processing) of verbal interference/facilitation in the response priming paradigm with gender categorization task. Participants in our study were asked to judge whether the presented stimulus was a female or male face that was briefly preceded by a gender word either congruent (prime: 'man', target: 'man'), incongruent (prime: 'woman', target: 'man') or neutral (prime: 'day', target: 'man') with respect to the face stimulus. We investigated whether related word-picture pairs resulted in shorter reaction times in comparison to the pictures with neutral words (facilitation) and whether unrelated word-pictures pairs resulted in slower reaction times in comparison to pictures with neutral primes (interference). We further examined whether these effects (if any) map onto response conflict or aspects of target processing. In addition, we introduced identity ('man', 'woman') and associative ('tie', 'dress') primes in order to investigate the cognitive mechanisms of semantic (introduced by associations) vs. Stroop-like (introduced by identity words) effects in response priming. We analyzed responses and reaction times using drift diffusion model to examine the effect of facilitation and/or interference as a function of the prime type. We found that regardless of prime type words introduce facilitation effect, which maps to the processes of visual attention and response execution.

Introduction

Words facilitate visual decisions in a variety of tasks. For example, the brief presentation of a word before a semantically related picture has been shown to facilitate the processing of the picture in a number of tasks such as detection of motion direction (Meteyard et al., 2007), word-picture matching (Boutonnet & Lupyan, 2015), recognition of ambiguous "Mooney" images (Samaha, Boutonnet, Postle, & Lupyan, 2018) and familiarity judgments (Amado et al., 2018). On the other hand, a word presented right before the picture can also lead to semantic interference, i.e. when a semantically unrelated word interferes with the judgement of an immediately following target compared to when the word is semantically related. In terms of behavioral performance this effect translates to longer reaction times and/or more errors (Wentura & Degner, 2010) as it has been shown to be the case in a variety of language tasks such as Stroop task(s), word-picture matching, and spoken-to-written word matching (Campanella, Crescentini, Mussoni, & Skrap, 2013; Faria, Alves, & Charchat-Fichman, 2015; Jefferies, Patterson, & Ralph, 2008). While facilitation effects have been interpreted as cognitive (with the locus of their influence being rooted either in the lexico-semantic system (Francken et al., 2015) or in the visual system (Boutonnet & Lupyan, 2015) some of the interference effects have been associated with response processing. For example, in the Stroop task the increased response latencies for naming the ink color of a conflicting color word (word: red, ink: green) in comparison to the non-conflicting (congruent) one (word: red, ink: red) are interpreted as a result of response conflict as opposed to informational conflict (Cohen, Dunbar, & McClelland, 1990; Duncan-Johnson & Kopell, 1981). Whether semantic facilitation and interference effects operate at the level of target and/or response processing is yet not entirely clear.

A paradigm, which is structurally similar to the Stroop task but less studied in terms of cognitive processes, is the response priming paradigm (Wentura & Degner, 2010). While Stroop-like effects appear at the response processing stage, response priming paradigms may tap into both target and response processing. In semantic priming paradigms two stimuli are subsequently presented and participants are instructed to perform a task on the second stimulus (target), while the first stimulus (prime) is deemed to be non-relevant to the task (for example, participants might have to classify letter strings as words or non-words with prime-target pairs being either semantically related or unrelated). In response priming paradigms instead, primes are either congruent or incongruent with the response that has to be given to the target (*prime*: skirt, *target*: woman, *response*: woman, *task*: categorization). In the Stroop task interference occurs when the ink color maps onto one response and the semantic meaning of the incongruent word to the other. In other words, interference arises due to the conflict between the responses. In response priming, similarly, interference occurs when the semantic meaning of the prime maps to one response and the semantic meaning of the target maps to the other response leading to response conflict (De Houwer, Hermans, Rothermund, & Wentura, 2002; Klauer, & Musch, 2003). The fact that response priming and Stroop tasks are structurally similar does not exclude the possibility that interference effects could be explained by either one or both processes of response competition and target processing. It is therefore of theoretical importance to investigate semantic facilitation and interference effects in a response priming paradigm.

In this study, we focused precisely on this aspect and investigated in a response priming paradigm effects of semantic facilitation and interference as reflected by changes in behavioral performance. We further explored whether facilitation and/or interference can be accounted for by mechanisms of response execution and/or target processing, which we formalized using the drift-diffusion model (see details further). Participants in our study were asked to judge whether the presented (target) stimulus was a female or male face when the target was briefly preceded by a gender word (prime) either congruent (prime: 'man', target: 'man'), incongruent (prime: 'woman', target: 'man') or neutral (prime: 'day', target: 'man') with respect to the face stimulus. Participants were instructed to decide about the gender of the target while ignoring the prime. We then looked at whether related word-picture pairs resulted in shorter reaction times in comparison to neutral word-picture pairs (facilitation) and whether unrelated word-pictures pairs resulted in longer reaction times in comparison to neutral word-picture pairs (interference). We further examined whether these effects (if any) map onto response conflict or target processing. In order to shed light on the cognitive mechanisms underlying semantic facilitation and interference effects, we used a well-established approach from the cognitive modelling literature, the drift-diffusion model (DDM) (Ratcliff & McKoon, 2008a; Smith & Ratcliff, 2004; Usher & McClelland, 2001).

Drift-diffusion model

In the DDM approach, the process of making a decision about the gender of a face is described as the stochastic accumulation of sensory evidence over time towards one of two decision boundaries (male or female response, for instance). Once enough evidence is accumulated and one of the two decision boundaries is reached, the associated response is produced (for example, female). In the DDM, a total of four parameters describe the processing components underlying the decision making process (see Figure 1): the rate at which evidence accumulates over time (*drift rate, v*), the amount of evidence that is necessary to produce a response (*boundary separation, A*), an optional a-priori bias for a specific response (*bias, z*) and finally the time required to complete non-decision processes, such as motor preparation and/or stimulus encoding (*non-decision time,* T_{er}).

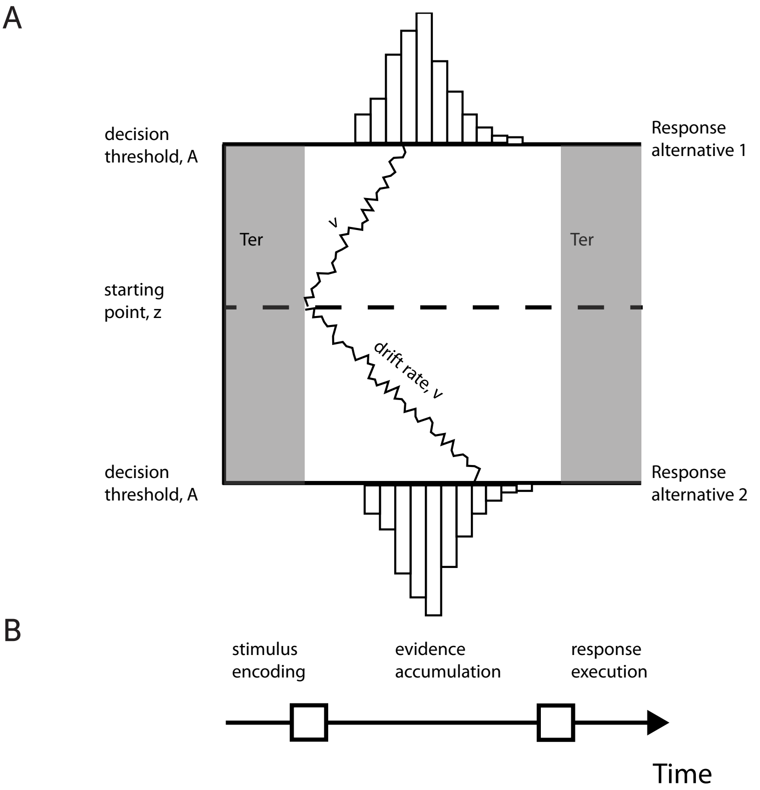


Figure 1: (A) The drift-diffusion model (DDM) with the four parameters: drift rate (*v*), boundary separation (*A*), starting point (*z*), non-decision component (*Ter*). **(B)** Stages of processing characterized by the DDM model. Different stages of processing are highlighted by the shadings in panel A indicating the mapping of the DDM components to the processing stages.

The DDM model has been successfully applied to choice reaction time data in various experimental tasks (see for reviews (Mulder, van Maanen, & Forstmann, 2014; Ratcliff & McKoon, 2008a) and the parameters recovered by the model have been shown to be well characterized in terms of cognitive processes (Voss, Rothermund, & Voss, 2004). The drift rate or the speed of evidence accumulation is "determined by the quality of information extracted from the stimulus" (Ratcliff &

McKoon, 2008b). For example in a color discrimination task, trials with less visible colors resulted in a lower drift rate as opposed to trials with the more visible colors (Voss et al., 2004). Crucially, the drift rate can be modulated by factors that are not only related to the stimulus being processed but also related to contextual information associated with the study episode. For example, in a memory recognition task, a word that was studied three times had a higher drift rate than a word that was studied only once (Ratcliff & McKoon, 2008b). Similarly, in an associative priming task, word primes that were associatively related to the target words resulted in a higher drift rate in comparison to word primes non-associatively related. It has been suggested that the drift rate in these contexts could represent "the quality of the match between a test word and memory" (Ratcliff & McKoon, 2008b). The accumulation of the drift rate stops when either one of the two decision boundaries has been reached. The amount of evidence that is needed to make a decision is characterized by the boundary separation, that is the distance of the boundaries from the starting point of the accumulation process, and it has been shown to be modulated by changes in task-strategy (e.g. response caution). For example, when participants were instructed to prioritize response accuracy over response speed, changes in behavioral performance due to the adoption of a new response criterion by the decision maker were explained in the DDM model in terms of a higher value for the boundary separation parameter. The higher boundary separation translates to a longer period of information accumulation and, as a result of it, fewer errors being made by the decision maker (Bogacz, Wagenmakers, Forstmann, & Nieuwenhuis, 2010). The starting point of the accumulation process instead reflects potential biases participants might have resulting in certain responses being "a priori more likely" (Mulder et al., 2014). For example, participants might a-priori favor a "word" response in a lexical decision task (Wagenmakers, Van Der Maas, & Grasman, 2007). In addition, it was also shown that in a color discrimination task, a higher reward for a certain response resulted in participants adopting a starting point (z) closer to the decision boundary for the response with the higher value reward (Voss, Rothermund, & Voss, 2004b). Finally, the model component that does not account for decision processes is referred to as non-decision time. It reflects either stimulus encoding (which may not be necessarily perceptual encoding, but access to memory in a memory task or lexical access in a lexical decision task, Ratcliff & McKoon, 2008b) or the time required to execute a motor response. These contributions are combined together in one parameter (T_{er}), which does not allow by itself the separation of encoding from response execution but rather allows for the separation of decision vs. non-decision processes. Studies that interpret T_{er} as encoding or execution parameter use auxiliary methods such as neuroimaging to facilitate interpretations. For example, in an fMRI study investigating age-related performance in a visual search task, changes in non-decision time were associated with targets accompanied by response-incompatible distractors in the elderly group. T_{er} was correlated with the FEF and dorsal fronto-parietal regions, which suggested a major contribution from the visual encoding process (Madden et al., 2019). An EEG study of figure-ground segregation instead found a correlation between N200 latency and the non-decision component suggesting that N200 tracks the completion of visual encoding (Nunez, Gosai, Vandekerckhove, & Srinivasan, 2019).

In the DDM model, the parameters are combined non-linearly to enable inferences on the complete distribution of reaction time data. Technically, this is done by computing for each accuracy interval or bin (e.g. 5 intervals of 20% accuracy performance increments) the relative RT distribution and then fitting a (Gaussian) random walk model to each of the quantiles of the RT distribution. An intuitive way to think about how the parameters of the DDM model combine to produce RT distributions is the following. As an example, assume a hypothetical classification task where subjects have to classify a face as female or male and we assign the female face response to the upper boundary (Response alternative 1 in Figure 1), and the male response to the lower boundary (Response alternative 2 in Figure 1). Assuming there is no bias for either of the two response and the same boundary separation for both male and female face targets (same amount of evidence to be accumulated) and no difference in non-decision time, a drift rate towards the female-response boundary (i.e. positive drift) would indicate faster correct responses for face-related judgments of female faces. By contrast, differences in threshold or non-decision time would suggest that overall RTs are either longer or shorter in female-related judgments compared to male-related judgments (depending on the directionality of the difference), regardless of the correctness of the response. Similarly, a higher boundary separation would predict on average slower responses (since more information has to be accumulated) and thus a longer mean RT for a particular condition would be predicted. Whether a response would be correct or not, however, would be mainly driven by the drift rate. “Mainly” is used here because boundary separation (A), drift rate (v) and bias (z) altogether contribute to the generation of choice RT data, specifically to the differentiation of fast and slow responses in both correct and incorrect responses. For comprehension purposes we point the reader to a simple qualitative taxonomy of the type of responses predicted by the DDM given a particular combination of A (boundary separation) and v (drift rate) parameters. If A is high and v is towards the correct response boundary (i.e. positive for upper boundary and negative for lower boundary), then the responses are predicted to be on average slow and correct. If instead A is high but v is towards the incorrect response boundary, then the responses are predicted to be on average slow and incorrect. If there is a change in the boundary separation parameter and, for example, if A is low and *v* is towards the correct response boundary, then the model predicts on average fast correct responses. If instead *A* is low and *v* is towards the incorrect response boundary, then the model predicts on average fast errors.

To sum up, in this study we capitalize on the DDM model since it provides an ideal analytical tool for disentangling in the response priming task the cognitive processes involved in semantic facilitation and interference. Importantly, previous drift-diffusion studies revealed intriguing cognitive mechanisms underlying behavioural performance in language-related tasks that use the response priming paradigm. In the primed Stroop task words-distractors were presented ahead of colored symbols that had to be categorized and the facilitation effect was explained in terms of a change in boundary separation among congruent and incongruent conditions (Kinoshita, de Wit, Aji, & Norris, 2017). This result however could be explained in terms of stimuli probability since a word prime was predictive of its matching response color (see p.833 (Kinoshita et al., 2017)). It has also been shown that predictive cue information results in boundary separation modulation (not drift rate) in a task where participants have to identify a house or face masked by noise, preceded by either a house or face cue with different degrees of reliability (Dunovan, Tremel, & Wheeler, 2014). The study of Kinoshita et al., is one of the recent studies on response priming that investigates language influences on target and response processing from the computational modelling perspective. Results are however not conclusive due to the predictability confounds described above. In this study language primes were not predictive of the upcoming target which allowed us to investigate the influence of language on facilitation and interference without introducing probabilistic confounds. Another interesting property that seem to affect participants' performance in the response priming task, and which is crucial for this study, is the type of semantic relationship between prime and target, which will be discussed further in the following section.

Type of semantic relationship

The type of semantic relationship between the target and the prime is another variable to be considered when investigating priming effects. For example, in the response priming paradigm, it was shown that associative and categorical primes involve different cognitive processes (Voss, Rothermund, Gast, & Wentura, 2013). While primes that belonged to the same category as the targets (i.e. categorical primes such as prime: lion, target: tiger) mapped onto the response execution stage, primes that were semantically associated with the targets (i.e. associative primes such as prime: king, target: crown) mapped onto the target processing stage. Specifically, the drift-diffusion analysis revealed that associative priming effects mapped onto the drift rate parameter, which indicated increased informational uptake for associative (prime: king, target: crown) word pairs in comparison to categorical ones (prime: lion, target: tiger). Furthermore, categorical primes mapped onto the non-decisional component, with congruent word-target pairs leading to a facilitation of the non-decisional processes, and incongruent ones resulting in interference. The authors explained categorical priming effects in terms of response competition, and associative priming effects in terms of spreading activation (Collins & Loftus, 1975). Crucially, categorical congruency effects were associated with response competition processes regardless of the relevance of the congruency dimensions (i.e., whether the task was lexical decision or semantic categorization did not affect the results). On the contrary, Gomez and colleagues (Gomez, Perea, & Ratcliff, 2013) showed that categorical prime-target pairs mediated both the non-decision and the drift rate components. In the study of Gomez, however, the authors used a variation of categorical primes involving identity primes (prime: house, target: house) instead of different words being related to each other categorically (such as prime: lion, target: tiger). Furthermore, they used a lexical decision task instead of the semantic categorization task used by Voss et al. which altogether might have led to differences in the experimental results. Other evidence from the word production literature shows a dissociation between associative and identity primes. For example, in picture-word interference, where participants name a picture and ignore a distractor word, picture naming is slower when the target image and distractor word are related in comparison to when they are unrelated (Glaser & Düngelhoff, 1984; Piai & Knight, 2018; Piai, Roelofs, Acheson, & Takashima, 2013b). Interestingly, other types of semantic relations such as for example associations, hypernym-hyponym, part-whole or nouns-verbs (Kuipers & La Heij, 2008; Lupker, 1979; Mahon, Costa, Peterson, Vargas, & Caramazza, 2007) result in facilitation or no modulation. Brain imaging studies have also shown a dissociation between associative and categorical relationships in terms of their neural bases (for a review see Mirman, Landrigan, & Britt, 2017). In sum, existing experimental evidence suggests that associations and categorical relations might have a differential effect on facilitation and interference, and therefore should be properly accounted for in tasks that involve response and attention control. In this study we specifically address the question of whether prime-target pairs that are related to each other either categorically (e.g., prime: 'man', target: man) or associatively (e.g., prime: 'tie', target: man) result in facilitation or interference effects and, if so, whether such effects occur at the level of response processing, target processing or both.

The present study

Our primary interest was verbal interference/facilitation in the response priming paradigm, either at the level of response execution or of target processing, when manipulating the type of semantic relationship linking prime and target. The facial features of the target pictures were morphed from male to female parametrically and the task included ambiguous faces sorely for participants' engagement purposes. The primes were either associations ('tie', 'dress') or identity primes ('man', 'woman'). In addition, primes were either congruent (prime: 'man', target: man), incongruent (prime: 'woman', target: man) or neutral (prime: 'day', target: man) with respect to the face. Participants were asked to decide about the gender of the target, while ignoring the prime. Participants' behavioral performance (responses and reaction times) was analyzed with the DDM approach to examine the effect of facilitation and/or interference as a function of the prime type.

The DDM approach focused specifically on testing two hypotheses. First we investigated whether congruency effects in response priming tap into the cognitive mechanisms associated with the processing of the response and/or target. Under the response competition account, congruency effects (e.g., the prime maps onto a "female" response and the face is a female face) would map onto the speed of the non-decisional processes (i.e., slower for incongruent and faster for congruent). Under the target processing account, congruency effects would map onto the speed of processing of the target picture (drift rate). In this case we expected the drift rate (*v*) to increase in congruent vs. incongruent word-picture pairs. Second, we investigated whether the type of semantic relationship - associative or identity – would have an influence on the direction of the effect, i.e., facilitation or interference. For this purpose, we tested whether identity words ("man", "woman") and associative words ("beard", "dress") tap into response conflict processes (T_{er}) and/or perceptual uptake (*v*). We furthermore tested whether identity words would enhance the rate of evidence accumulation as opposed to associative words and whether identity words would lead to faster motor preparation/execution in comparison to the associative ones.

To sum up, with this work we aimed to investigate whether associative and identity words relating to gender categories affect cognitive processing, specifically the stages of response execution and/or information evaluation of the target during gender categorization.

Methods

Subjects

The study was approved by the local ethics committee (CMO Arnhem-Nijmegen, Radboud University Medical Center) under the general ethics approval. All participants provided written informed consent approved by the Radboud University, Nijmegen. 47 volunteers (23 females) recruited via the Radboud Research Participation System (native Dutch speakers, right-handed, age range: 19 - 35 years, mean age = 24.74, *sd* = 3.53) took part in the study. We performed two additional studies to develop and pre-test the materials. All participants reported no neurological disease, and had normal or corrected-to-normal vision. All of the participants received monetary compensation for their participation.

Target pictures

A set of realistic 3D gender-morphed faces was created with the use of FaceGen Modeller 3.5 (Singular Inversions). The technical details of the computation method used by the software are discussed in (Blanz & Vetter, 1999a). We created 81 Western face identities, for which we gradually morphed gender features from extremely female to extremely male in 10 equal steps (Figure 2, A). Face stimuli were cropped to remove hair and ears and presented frontally. We controlled for luminance using the SHINE toolbox for MATLAB (Willenbockel et al., 2010).

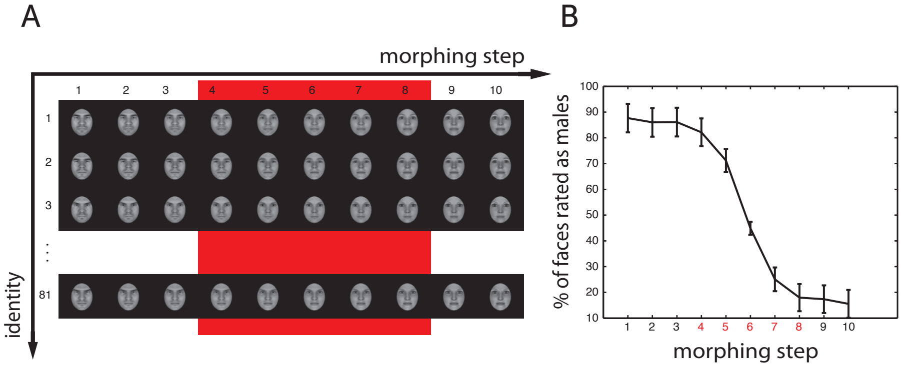


Figure 2: (A) Stimuli set with 81 artificially generated identities each of which was morphed in ten steps between male and female spectrum. The red square indicates the stimuli selected for the experiment. **(B)** Proportion of faces evaluated as males as a function of morphing step. Faces of morphing steps 4 (very male), 5 (less male), 6 (ambiguous), 7 (less female), and 8 (very female) were selected for the main experiment. Faces of step 6 were not included in the analysis (white marker in Figure 2 A and black in Figure 2 B).

Target pictures evaluation and selection

In order to obtain subjective perceptual ratings of the face gender, we conducted a stimulus evaluation experiment where subjects performed a forced two-choice task on the gender of each of the faces of the experimental set. A separate pool of 47 volunteers (24 males, native Dutch speakers, age range 20 - 32 years, mean age = 23.83, *sd* = 2.88) participated in a separate experiment to evaluate the face stimuli. Another pool of volunteers (24 females, native Dutch speakers, right-handed, age range 20 - 53 years, mean age = 26.75, *sd* = 6.65) performed the semantic ratings of priming words. Each trial in the evaluation experiment started with a fixation cross that stayed on the screen for 1500 ms, after which the target was visually presented for 500 ms. Participants were able to deliver a response for an additional 2000 ms after the picture was removed from the screen. Based on the responses, we identified the morphing step of the faces that was perceived ambiguously. We included five morphing steps in the main experiment: the most ambiguous face in the middle of the continuum, plus two steps in either direction away from the middle point (Figure 2, B). We define faces of morphing step 6 as ambiguous faces, and all the others as less ambiguous, with morphing step 4 as the most male face and morphing step 8 as most female face. The percentage of faces evaluated as males for ambiguous faces (step 6 in the Figure 1, B) was 47.87 % (*sd* = 16.45), for the selected extremely male faces (step 4), 84.21 % (*sd* = 27.19), and for selected extreme-female faces (step 8), 16.80 % (*sd* = 25.57).

Prime words evaluation and selection

The experiment consisted of two conditions. In the first (*identity*) condition, the labels “man” (English: *man*) and “vrouw” (English: *woman*) were used as primes. In the second (*associative*) condition, a set of gender-associated words such as “mascara” and “tie” were used as primes (see Appendix and Table 2). The words for this condition were preselected using the database from the “small world of words project” (De Deyne, Navarro, & Storms, 2013a), and subsequently rated by naïve participants. In the rating experiment, the participants had to indicate for each word how related that word was to the words “male” or “female” on a 7-point scale (-3 = related to male; + 3 = related to female). For half of the participants, female and male axes were swapped. Based on the rating outcomes, a selection of 40 words was made, which included 20 male-related and 20 female-related words. In addition, 20 words from the semantic category “furniture” (the neutral condition) and 20 catch words from diverse semantic categories were selected. The furniture and filler words were associated with neither the male nor the female categories according to the ratings. Male and female-associated words were matched for word length, frequency per million and concreteness (all $p > .30$). The catch trials were excluded from subsequent analyses. Frequencies for all words were extracted using the Subtlex corpus (Keuleers, Brysbaert, & New, 2010). Mean frequency, concreteness, and length of the materials used are indicated in Table 2.

Table 2 Characteristics of the words used in the priming experiment.

Associated groups	Length	Frequency	Concreteness
female (N = 20)	M = 6 (SD = 2.1)	M = 60.4 (SD = 101.5)	M = 4.1 (SD = 0.6)
male (N = 20)	M = 5.8 (SD = 1.9)	M = 108.2 (SD = 204.6)	M = 4.3 (SD = 0.6)
neutral (N = 20)	M = 5.8 (SD = 1.9)	M = 69.6 (SD = 187.5)	M = 4.1 (SD = 0.9)

Semantic Similarity

It could be argued that the proposed associative words in the stimuli list we used contain both associations (‘tie’) and identity-like words (‘brother’, ‘father’). Therefore, we sorted the initial associative set into associations and labels (see Suppl. Table 1.2). Further, to control for the semantic similarity between the prime words (identity: ‘man’, labels: ‘father’ , associative: ‘tie’ , ‘dress’) and target concept (‘male’, introduced by a male face, or ‘female’, introduced by a female face), we used the *snaut* tool (Mandera, Keuleers, & Brysbaert, 2017) which is based on word2vec representations (Mikolov, Chen, Corrado, & Dean, 2013). The word2vec model represents words’ semantics as a vector of features and the semantics of a certain word can be characterized by comparing the vector representations. The measure of semantic similarity we report here is cosine similarity, which has particular advantages over other measures such as Euclidean and Manhattan in cases where the vector magnitude matters. First, we calculated the semantic distance between each of the primes to the target concepts (associations: ‘beard’ – ‘man’, labels: ‘father’ – ‘man’, identity words: ‘man’ - ‘man’) in terms of cosine distance (see Suppl. Table 1.2). Next, we tested whether the proposed word sets (associations, labels, identity words) differed in semantic measures using Bayesian ANOVA which accounts for the non-equal number of words per word group. The null hypothesis states that there is no difference between the conditions of interest whereas the alternative posits that the conditions of interest are different. A Bayes factor (BF) is defined as the ratio between the evidence in favor of the alternative hypothesis (H_1) over the evidence in favor of the null hypothesis (H_0), denoted by the subscript 10 in the Bayes factor abbreviation $BF_{10.}$ BFs estimate graded evidence in favor or against the alternative hypothesis (Wagenmakers et al., 2018) and can be interpreted as follows: BF_{10} = 1-3 indicates “anecdotal” evidence for H_1 compared to H_0; BF_{10} = 3-10 indicates “moderate” evidence for H_1 compared to $H_{0;}$ BF_{10} = 10-30 indicates “strong” evidence for H_1 compared to $H_{0;}$ BF_{10} = 30-100 indicates “very strong” evidence for H_1 compared to $H_{0;}$ BF_{10} > 100 indicates "extreme" evidence for H_1 compared to H_0. Bayesian ANOVA was carried out using JASP (JASP Team, 2018).

The results of the analysis are presented in the Figure 3. Words of different types, fully overlapping with the picture ('man', 'vrouw', i.e., identity), less overlapping ('father', 'sister', i.e., labels) and associative items ('tie', 'dress'), translated to different semantic distances from the target concept (main effect of semantic distance, BF_{10} = 1.29 x 10^{47}). Specifically, identity words ('man/vrouw') had higher similarity (lower cosine distance) with the target concept ('man', 'vrouw') in comparison to the partially overlapping words (e.g. 'father', 'sister'), BF_{10} = 3.70 x 10^{6}. Associative words were further from the target concept in comparison to the identity words, BF_{10} = 2.54 x 10^{22}) whereas labels were closer to the target concept in comparison to the associative words, BF_{10} = 250.437.

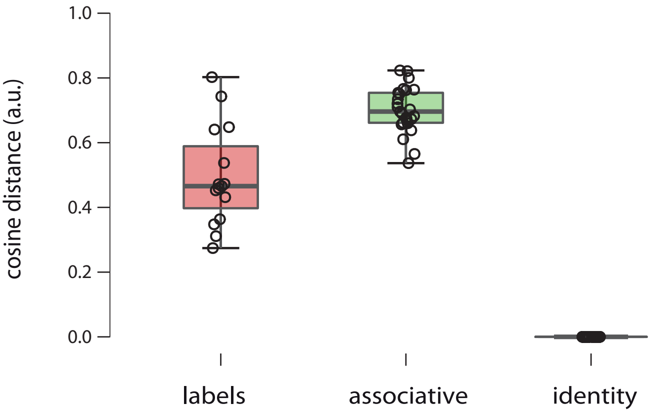


Figure 3: Semantic distance between prime words (labels, associations and identity words) and target concept introduced by the face ("female"/ "male") color coded for type of prime word: red, labels ("father" / "daughter"); green, associations ("beard" / "skirt"); black, identity words ("man" / "woman").

The semantic distance between identity words and labels was larger (evidenced by a "very strong" Bayes factor) than between labels and associative words (labels - associations > labels - identity words: BF_{10} = 2.548 x 10^{22}).

Therefore, for the first set of analyses (see Congruency Analysis), we collapsed labels and associations (the non-identity condition), especially in light of the fact that in the current paper we did not focus on the various types of similarity between

the words but rather focused on the contrast between identity words and all other words that point to the same concept but have larger semantic distances.
However, since the difference between the labels and associations is statistically significant, we provide additional analyses excluding the labelling primes and thus only comparing associative and identity primes. These additional analyses are explained further below (see section Behavioural analysis, Secondary Congruency Analysis: associative versus identity primes).

Procedure

Participants were seated in a comfortable chair in front of a computer screen in a sound-protected room. On each trial, a prime word was presented for 250 ms, after which a fixation cross remained on the screen for 300 ms. The visual target was presented for 500 ms followed by a jittered inter-trial interval of 1500-3000 ms (Figure 4). The total number of trials was 800. The experiment was divided into 2 blocks (400 trials per each block) with an optional break in-between the blocks. Overall, for the associative prime condition, there were 20 male-, 20 female-related and 20 neutral words. For identity words, we had the word ‘woman’/’man’ presented 20 times each, and 20 neutral words (the same that we used in associative condition). The words were presented with each morphing step (face identity was shuffled with no repetition) and each prime word was repeated 5 times. The trial order was randomized. The total duration of the experiment was approximately 90 minutes.

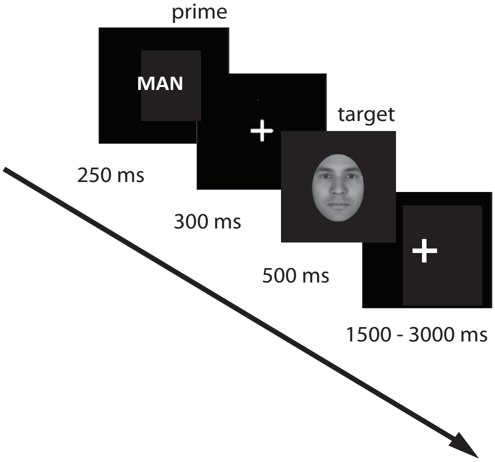


Figure 4: Experimental design. A prime (label or word associated with the target concept) was presented for 250 ms, followed by a fixation cross (300 ms), after which the target picture was presented followed up by a jittered fixation cross of 1500-3000 ms. Participants had to decide about the gender of the presented face by button press.

Task

Participants were instructed to decide upon the gender (male or female) of the face based on the image presented, and to respond with a keyboard button press (middle/index finger; the mapping of the response buttons was counterbalanced across participants). Participants had up to 2 seconds to respond after onset of the picture and were instructed to skip the trials on which the prime belonged to the category "furniture". This "go/no-go" task ensured that participants read the prime word.

Analyses

Behavioral analysis

We investigated priming effects by separately analyzing measures of behavioral performance: reaction times (RT) and choice responses. We performed a 2 by 3 repeated measures analysis of variance (rm ANOVA) with *prime type (identity* or *associations)* and *congruency (congruent, incongruent, neutral)* as factors. We preselected congruent (target: male face, prime "man"; target: female face, prime "female"), incongruent (target: male face, prime "female"; target: female face, prime "man") and neutral (target: male face, prime "day"; target: female face, prime "day") word-target pairs collapsing across very- and less-gendered morphing steps (step 4, 5; step 7, 8). We performed post-hoc comparisons for main effects using Holm correction. Morphing step 6 was excluded from all analyses since ambiguous faces can require a different configuration of the decision process in comparison to the male and female gendered faces. Whereas faces of step 6 are marked by high-uncertainty, faces of steps 4,5,7,8 are instead marked by low uncertainty. Given this potential difference and given our focus on the potential differences among types of word primes, we decided to exclude step 6 items to avoid potential confounds in terms of uncertainty. It could be argued that the associative condition introduced in this study may be viewed as a conjunction of "labels" and "associations", which would make the comparison between the two groups confounded (see "Semantic similarity" above in the materials section). Thus, in a secondary set of analyses, we repeated a 2 by 3 rm ANOVA analysis on the associative words, omitting potentially confounded words in the non-identity condition (i.e., the 15 "label primes" (e.g. father) and resulting in 25 associative words, see Figure 3 and Supp. Table 1.2) using the Bayesian approach, which accounts for unequal number of trials.

Moreover, it could be argued that facilitation/interference effects can be modulated by repetition of the materials (i.e., the identity primes are repeated more than associative primes), therefore we also performed an analysis of repetitions. First, we calculated the number of repetitions for the conditions of interest (congruent, incongruent, neutral) separately for associative and identity words. Then we calculated the average effect of interference (unrelated - neutral) and facilitation (related - neutral) for associations and identity words. Since the number of repetitions was unequal per condition for identity words (~ 80 repetitions) and for associative words (~ 50 repetitions), we calculated the facilitation/interference effect across the whole session with a step augmentation of 10 trials in order to better illustrate the differences between the identity/associative conditions,. Further, we performed four ANOVAs (separately for interference and facilitation per associations and identity words) with number of trial repetitions as dependent variable. The analyses were performed using JASP (JASP Team, 2018).

Hierarchical Drift-diffusion Model

In order to gain insights into the processing components underlying categorization in the identity and associative conditions, we analyzed choice reaction time data with the hierarchical drift-diffusion model. The analysis was implemented in the Python toolbox HDDM 0.6.0 (Wiecki et al. 2013). One of the main advantages of the hierarchical Bayesian framework is that the simultaneous estimation of the model parameters at both the single subject and group levels enhances statistical power since fewer trials are required to recover the parameters and the estimates are less susceptible to outliers (Wiecki, Sofer, & Frank, 2013), making it an appropriate analytic approach for the present study. Models with different combinations of free parameters were fit to the data via Markov Chain Monte Carlo (MCMC) fitting routines. We used an accuracy coding scheme for the responses for congruent/incongruent/neutral prime-target pairs with the upper boundary reflecting a correct face categorization and the lower boundary an incorrect one. We defined the model space by allowing the parameters to vary freely over the factors of interest (*congruency, prime type*) of the experimental design (see Suppl. Table 2).

For each model we evaluated the rate of convergence of the numerical fitting routines and then the ability of the model to capture the observed RT distributions. Models that failed to reach convergence or failed to capture the observed RT distributions were excluded from further analyses. Finally, the remaining models were compared against each other by computing the relative Deviance Information Criterion (DIC), which is a measure of the goodness of the model fit to the data that penalizes for the complexity of the model (Schwarz, 1978). A rm ANOVA was then used to test for significant differences in the parameter estimates of the best fitting model and to quantify the evidence in support of a given hypothesis.

Results

Behavioral analyses

Congruency Analysis

Behavioral results are summarized in Figure 5. We highlight that the associative condition consists of both labels ("father" / "daughter") and "pure" associations ("beard" / "skirt"). For rationale see Methods, *Semantic Similarity*. Participants' speed of response varied as a function of congruency (main effect of *congruency*: *F* (2, 92) = 7.63, *p* < .001). On average, congruent words resulted in faster RTs in comparison to incongruent ones (congruent > incongruent: t = -6.13, p < .001, SE = 0.004, Mean difference = -.027) and to neutral ones (congruent > neutral: t = -4.86, p < .001, SE = 0.004, Mean difference = -.021). We also found that the incongruent words resulted in slower RTs in comparison to neutral ones (incongruent > neutral: t = 2.56, p = .04, SE = 0.002, Mean difference = .006).

We also found that participants' performance varied depending on the type of the prime (main effect of *prime type*: *F* (1, 46) = 4.70, *p* = .035). On average, identity words resulted in longer RTs in comparison to associative words: (associative > identity: t = -2.16, p = .035, SE = 0.007, Mean difference = -.014).

Congruent pairs with associative and identity words resulted in shorter RTs in comparison to neutral pairs (congruent > neutral (identity): t (46) = -3.53, p < .001; congruent > neutral (associations): t (46) = -5.07, p < .001). Interestingly, only incongruent pairs with labels resulted in longer RTs in comparison to the pairs with neutral words (incongruent > neutral (identity): t (46) = 2.32, p = .024; incongruent > neutral (associations): t (46) = 0.40, p = .689).

For the proportion of correct responses, we did not find a difference for *prime type* (identity/associations): *F* (1, 46) = .014, *p* = .90. The effect of congruency did reach significance: *F* (2, 92) = 23.48, *p* < .001. We did not find an interaction between *prime type* and *congruency*: F (2, 92) = 0.13, p = .87.

Secondary Congruency Analysis

We further excluded labels ("father" / "daughter") from the associations ("beard" / "skirt") and repeated the analysis in the Congruency Analysis. For the RT, we found a main effect of *congruency*: *F* (2, 92) = 22.53, *p* < .001. On average, congruent prime-pairs resulted in faster RTs in comparison to incongruent word-pairs (t = -5.42, p < .001, SE = 0.004, Mean difference = -.023) and to neutral words (t = -4.63, p < .001, SE = 0.004, Mean difference = -.019). There was no difference between unrelated and neutral word-target pairs: t = 1.78, p = .081, SE = 0.002, Mean difference = .004.

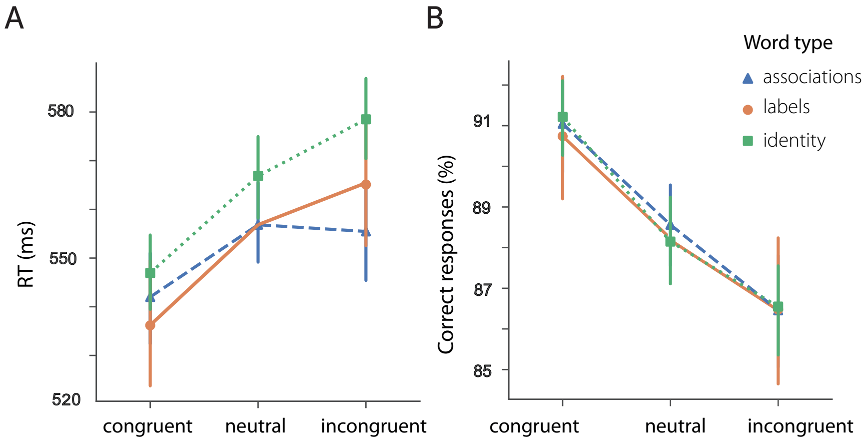


Figure 5: Mean reaction times **(A)** and correct responses **(B)** for identity ("man" / "woman") and associations (mix or associations ("beard" / "skirt") and labels ("father" / "daughter") as a function of semantic relatedness. Error bars represent 95 % confidence interval (CI); RT = response time.

We also found that there was a main effect of *prime type*: *F* (1, 46) = 5.10, *p* = .029. On average, associations resulted in shorter RTs in comparison to identity words: t = -2.25, p = 0.029, SE = 0.006, Mean difference = -.014.

We also found an interaction between *congruency* and *prime type*: *F* (2, 92) = 5.45, *p* = .006. To investigate the interaction we performed a series of pair-wise t-tests with the following results: both related associative and related identity primes resulted in shorter RTs in comparison to the neutral primes (congruent > neutral for associations: t (46) = -3.959, p < .001; congruent > neutral for identity: t (46) = -3.537, p < .001). As for the incongruent condition, only identity pairs resulted in interference (incongruent > neutral for identity: t (46) = 2.328, p = .024; incongruent > neutral for associative: t (46) = -0.82, p = .413).

For the proportion of correct responses we found a main effect of *congruency*: *F* (2, 92) = 19.58, *p* < .001. We did not find a main effect of *prime type*: *F* (1, 46) = .002, *p* = .96, nor an interaction effect between *congruency* and *prime type*: *F* (2, 92) = .084, *p* = .92.

Analysis of repetitions

Since there was intrinsically a different number of trials for associative vs. identity words, we performed a repetition analysis with the purpose of investigating the facilitation and interference effects as a function of the number of trials. To reiterate, we defined the facilitation effect as the difference between congruent and neutral prime-target pairs. The interference effect was defined as the difference between incongruent and neutral pairs. There were on average 80 repetitions for identity primes (for each of the related and unrelated conditions) and 50 repetitions for associative primes (for each of the congruent and incongruent conditions) per subject.

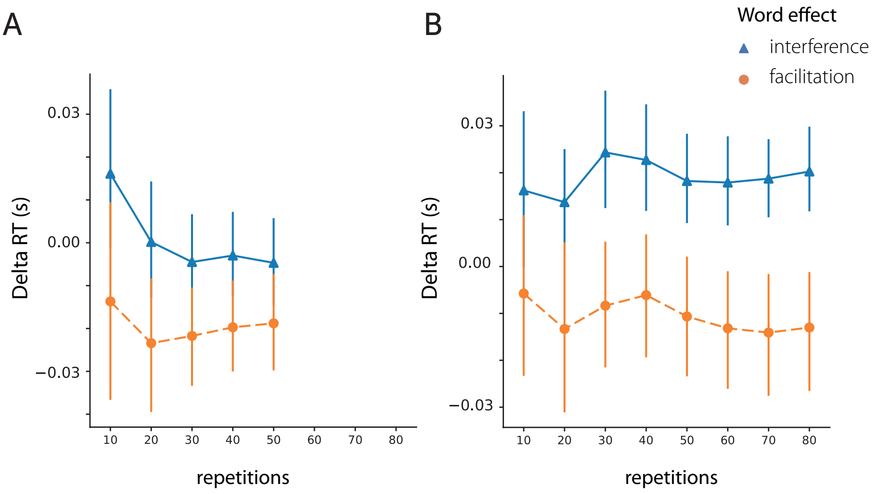


Figure 6: Interference and facilitation effects across trials for (A) associative words ("beard" / "skirt") and (B) identity ("man" / "woman") words. Error bars represent 95 % confidence interval (CI); RT = response time.

Within each of the conditions - associative or identity - we tested the effect of repetition separately for interference and facilitation effects. For the identity condition, we did not find a repetition effect neither for facilitation (*F* (7,322) = 0.96, *p* = .46) nor for interference (*F* (7,322) = 0.91, *p* = .49). For the associative condition, the repetition effect did not modulate the facilitation effect (*F* (4, 184) = 0.71, *p* = .58). However, the interference effect was affected by the repetition effect (*F* (4,184) =4.39, *p* = .002). This was mainly driven by a larger interference effect for the first 10 trials. Specifically, analysis of repeated contrasts showed significant difference only between 10 vs. 20 trials (t = 2.69, p = .008, SE = .006, Estimate = .016), but neither for 20 vs. 30 (t = 0.79, p = .42, SE = .006, Estimate = .005) nor for other repeated contrasts.

To sum up, the analysis investigating the effects of *congruency* and *prime type*, even when accounting for the potential confounds in the prime type, showed consistent results. Particularly, both identity words and associations resulted in a facilitation effect, but only identity words produced an interference effect. However, as we show in the Figure 6A, the repetition of items did affect the associative words, which reduced after 20 trials.

Hierarchical drift-diffusion modeling

Next we conducted a drift-diffusion analysis using RTs and choice responses from the associative and identity conditions (following the Secondary Congruency Analysis).

Model Convergence and Model Fit

For all of the analyses reported, the MCMC (Gelman & Rubin, 1992) fitting routines were run for 25.000 iterations with a burn-in period of 10.000 iterations and a thinning of 1. Model convergence was assessed by examination of the posterior samples and of the R-hat statistic, which is a measure of convergence among multiple MCMC chains (three for the present study). Posterior density estimates, which are stable over multiple samples, indicate that the fitting routines have converged to a fixed estimate. A R-hat statistic below 1.1 indicates that chains with different starting values have converged to the same posterior estimate. Successful convergence was confirmed by a MCMC error for all of the parameters smaller than 0.01. We further performed a comparison between observed and recovered RT distributions produced by the model (see Supplement, figure 1). After assessing convergence, we carried out a quantitative comparison of model-candidates by computing the associated DIC score for each model. There were 15 model-candidates (see Suppl. Table 2), which were constructed by setting three model parameters (A, Ter, v) as free of fixed along two experimental factors (*congruency* and *target*). DIC is a measure of the goodness of fit of the model to the data that is penalized for the complexity of the model and therefore a model with a lower DIC score is to be preferred over an alternative model with a higher DIC as the most parsimonious explanation of the data. Models that did not reach convergence were discarded and not included in the DIC comparisons.

Below we report modelling results using behavioural data from Secondary Congruency Analysis. The model that best described the data (i.e. the model with the lowest DIC score) was the model with the following parameters estimated per subject: drift rate (v) free over *congruency* and *prime type*, threshold (A) free over *prime type*, non-decision time (T_{er}) free to vary across *congruency* and *prime type* (see Supplement, table 2 for details). Conventionally, a DIC difference of more than 10 indicates that the evidence in favor of the winning model is substantial (Burnham, Anderson, & Burnham, 2002). Because the difference between the winning model (model 6, DIC -16601.7) and the second-best model (model 12, DIC -16588.0) exceeds 10 (13.7), we consider this evidence enough for selecting model 6 as the most parsimonious account of the data and therefore further analyses focus on this model.

Model parameter analysis

Congruency analysis for HDDM parameters

Results of the modelling analysis are summarized graphically in Figure 7.

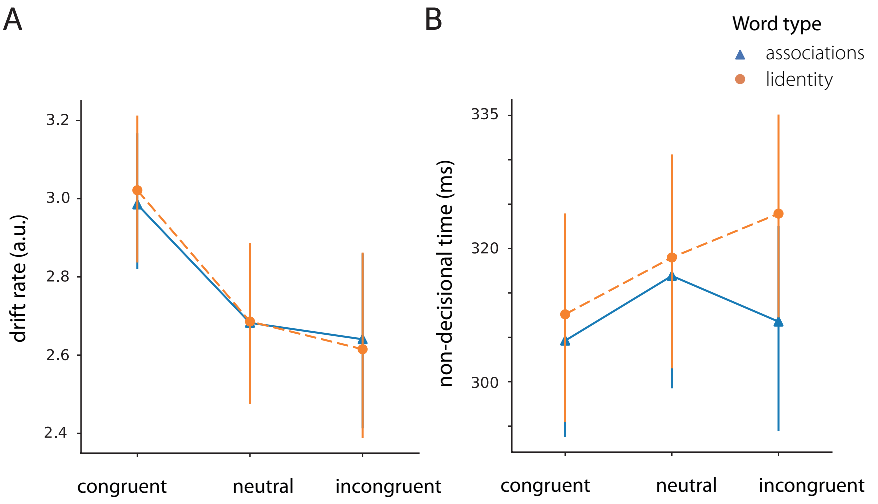


Figure 7: Congruency analysis. Posterior estimates of the hierarchical drift-diffusion model for the drift rate (*v*) and non-decisional parameter (*Ter*). Error bars represent 95 % confidence interval (CI).

Drift rate

We found a *congruency* effect (*F* (2, 92) = 27.34, *p* < .001). Particularly, congruent pairs had an increased drift rate in comparison to incongruent ones (t = 5.59, p < .001, SE = 0.067, Mean difference = .376) and in comparison to neutral ones (t = 6.88, p < .001, SE = 0.046, Mean difference =.0319). We did not find a difference between incongruent and neutral pairs (t = -1.168, p = .249, SE = 0.048, Mean difference = -.056).

Boundary Separation

We did not find a *congruency* effect (*F* (1, 46) = 2.63, *p* = .11).

Non-decision component

We found an effect of *congruency* (*F* (2, 92) = 5.81, *p* = .004). Congruent word-picture pairs had faster non-decisional time in comparison to incongruent (t = -2.78, p = .023, SE = 0.002, Mean difference = -.007) and neutral pairs (t = -2.92, p = .016, SE = 0.002, Mean difference = -.007). There was no difference in the non-decision time between incongruent and neutral pairs (t = -0.049, p = .96, SE = 0.002, Mean difference = -1.050e-4). We did not find either a main effect of *prime type* (*F* (1, 46) = 2.06, *p* = .15) or an interaction between *congruency* and *prime type* (*F* (2, 92) = 2.74, *p* = .069).

To sum up, we found a congruency effect for the drift rate (*v*) and the non-decisional parameter (T_{er}), but not for the decision boundary (*A*).

Discussion

In this study, we investigated whether the different types of semantic relationship (Mirman et al., 2017; Wentura & Degner, 2010) (associations or identity words) result in facilitation and/or interference in response priming of a gender categorization task. Participants had to decide about the gender of presented faces after having seen a word prime. From the analysis of reaction times, we found that both identity (e.g., prime "man") and associative (e.g., prime "beard"/"father") words resulted in a facilitation effect (congruent vs. neutral) whereas only identity words resulted in interference (incongruent > neutral). We further combined RTs and choice responses within the analytical framework of the DDM with the purpose of investigating the cognitive processes underlying facilitation and/or interference. We found a facilitation effect in both associative and identity words that translated to modulations in drift rate and non-decisional time.

Congruency with the target words facilitates information processing of the target picture

Words are one of the top-down factors (such as reward or task strategy) that influence perceptual decisions. Indeed, it has been shown that a larger reward for one of the response options or increased likelihood of occurrence of two events results in an enhanced starting point of evidence accumulation for that particular response (Mulder, Wagenmakers, Ratcliff, Boekel, & Forstmann, 2012). It has been recently proposed that language affects perception by setting predictive priors that sharpen perceptual representations (Simanova, Francken, de Lange, & Bekkering, 2016). Kinoshita et al. (2017) indeed showed that the brief presentation of a color word followed by the presentation of a color sign to be categorized resulted in a facilitation effect which translated to boundary separation and starting point modulations. This prompted the idea that words indeed affect perceptual decisions in a predictive fashion. However, this could be due attributed to the internal statistics of the experiment - words were predictive of the upcoming color of the target. In our experiment we unambiguously show that when words are not-predictive of upcoming features of the target stimuli, neither modulations in threshold nor starting point are manifested. Instead we found that words result in an increased drift rate which we interpret in terms of faster target processing speed.

Our finding supports well-established facilitation effects of language on perceptual decisions. For example, it has been shown that language can speed up recognition of visually presented objects (perceptual sensitivity or d′ prime) as has been demonstrated for the identification of facial expression (Carroll & Young, 2005) and for the detection of motion direction (Meteyard et al., 2007). Effects of language on visual perception have been demonstrated across different tasks and perceptual domains, including color categorization (Gilbert, Regier, Kay, & Ivry, 2006, 2008; Winawer et al., 2007) and face recognition (Anderson, Serences, Vogel, & Awh, 2014; Landau, Aziz-Zadeh, & Ivry, 2010). Experimental studies in the language domain are prone to interpret perceptual sensitivity results in terms of perceptual advantage, i.e. language tapping in the low-level representations (Meteyard et al., 2007). However, the theoretical premises of the drift rate describe it as a post-encoding measure which does not reflect the low-level encoding (of a target picture), but rather reflects an intermediate processing stage between stimulus encoding and response execution.

Usually in priming studies both prime and target are words and the modulations in drift rate are therefore interpreted in terms of increased spreading activation of a lexico-semantic nature (Voss, Rothermund, et al., 2013). However, in cross-domain priming (prime: word, target: picture) properties of the visual stimulus might change the nature of the process reflected by the drift rate. For example, it has been shown that the physical strength of the stimuli (i.e., intensity) was captured by a late event-related EEG potential, the centro-parietal positivity (CPP), which also tracked subjective perceptual experience (above the physically presented evidence) (Tagliabue et al., 2019). Another study showed that the rate of evidence accumulation is correlated with P300 which scaled with target detection difficulty (Twomey, Murphy, Kelly, & O’Connell, 2015) and indexed the duration of stimulus evaluation processes (Duncan-Johnson, 1981; Kutas, McCarthy, & Donchin, 1977).

Altogether these studies suggest that the modulation of evidence accumulation can reflect a separate meta-process. This could be tested by looking at whether the advantage in processing speed is due to lexico-semantic, visual or meta-processing facilitation, for example using EEG. One would expect a modulation of either the N400 (see for review (Kutas & Federmeier, 2011) in case of a semantic advantage, a P300 / CPP modulation in case of a attention or P1 in case of an early visual (Boutonnet & Lupyan, 2015) advantage. Regardless of the exact nature of the drift modulation, we show that the priming effects modulate informational processing of the picture. To clarify which type of information is needed and, as a consequence, is reflected in the drift rate requires a combination of mathematical modeling and neuroimaging tools, which may be of use for future studies.

On the processing of associations and identity words

We found that associative and identity words differ in the magnitude of RTs: longer for identity words in comparison to associative ones (no difference was found in accuracy data). It is known that upon repeated presentations of an item the chance of an error can be both diminished (repetition priming) and increased, resulting in cumulative semantic interference (Oppenheim, Dell, & Schwartz, 2007). For example, in a continuous naming paradigm subjects name pictures that belong to different categories and the naming times increase linearly with the number of pictures belonging to that category. Interestingly, the repetition of an item produced the same cumulative interference effect as additional novel exemplars in the category (Navarrete, Mahon, & Caramazza, 2010). We show that in a semantic categorization task with cross-domain priming the repetition of an identity prime-target resulted in longer RTs in comparison to the repetition of an associative prime-target pair, which is suggestive of cumulative semantic interference. To highlight, it can be that associative memories do not lead to cumulative semantic interference because categories and associations are represented differently as systems. Therefore identity prime-target items, which semantically belong to the same category (word: "man", picture: man), result in cumulative semantic interference, and associative prime-target pairs do not produce cumulative semantic interference because they point to different categories (word: "beard", picture: man). This raises questions about the non-semantic nature of associations and in general about the differences between the organization of semantic and associative memory systems (De Deyne, Navarro, & Storms, 2013b), which future studies should address.

In spite of showing the differential effect in RTs for association and identity words, we showed that this difference cannot be attributed to the speed of evidence accumulation. Traditionally, semantic priming effects have been explained in terms of memory aliasing, a process that helps integrating contextual linguistic information from the prime with the visual target – see the spreading-activation theory of semantic processing (Collins & Loftus, 1975) and the compound-cue account (Ratcliff et al., 1988). According to the spreading-activation theory, semantic memory can be seen as a network of interconnected nodes. If two nodes share semantic features, they are connected and the semantic distance determines the strength of this connection. This theory predicts that identity primes (i.e. the words "man" and "woman") would lead to a greater accessibility of the target in memory in comparison to the associative primes. Here, however, we found that the drift rate does not change as a function of semantic distance, which indicates that the drift rate does not necessarily reflect lexico-semantic memory effects but rather metacognitive processes. Previous experimental evidence suggests that briefly presented primes influence behavior via meta-cognitive fluency heuristics (Whittlesea & Williams, 1998, 2000). Fluency, or the meta-cognitive experience of the ease with which we process information, affects a wide variety of decisions (categorization: Oppenheimer & Frank, 2008; familiarity: Monin, 2003; lexical decisions: Potter, Donkin, & Huber, 2018). In a general sense, decisions can be made not only on the basis of the content but also on the basis of the feeling of how easy it is to make a decision – in this sense fluency operates as a heuristics that facilitates decision making (Schwarz, 2004).

In summary, we show that both identity and associative words increase the processing speed of the visual target. The fact that processing speed does not reflect differences in RT attributable to a semantic cumulative effect suggests that drift rate does not reflect in this case processes of a lexical-semantic nature.

Cognitive processes related to the response priming paradigm

Previous studies in the field of language investigated whether semantic effects in response priming paradigms (e.g., the words "A" or "B" are used as primes and the participants have to make an "A" or "B" decision on the target) can be explained by response facilitation at the motor stage and/or target processing facilitation (Wentura, Voss et al., 2013). In other words, it has been proposed that a prime can pre-activate a certain motor response option bypassing target processing (see the interference-facilitation account of linguistic priming effects in De Houwer, Hermans, Rothermund, & Wentura, 2002; Klauer, & Musch, 2003). In this view the prime can either facilitate motor execution if the upcoming target is coherent with the prime or interfere with the execution if there is a prime-target mismatch. In the current study we found that both the speed of evidence accumulation and non-decisional time were affected by the primes. This suggests that the primes do not by-pass the evaluation of the target, but rather exert their influence on the target and thus ruling out the interference-facilitation account. In conclusion, in this study we found that both associations and identity words in response priming led to a facilitation of face gender categorization. This effect was mapped to both response and target processing: when related to the target both prime types resulted in increased processing speed and faster motor response preparation. This result highlights the multidimensionality of the cognitive processes affected by language.

3

When task matters: the drift diffusion perspective on language-guided face gender categorization

Abstract

Language influences visual perception, for example, by facilitating categorical decision making about visual stimuli (Carroll & Young, 2005; Francken et al., 2015; Meteyard et al., 2007). Multiple studies have shown that language can affect stimuli-specific representations (Brouwer & Heeger, 2013; Landau et al., 2010), suggesting that language augments visual perception at a rather early stage. However, the mechanism of this facilitation is still not clear. Here we investigated whether the engagement of stimulus-specific representations can reflect the speed of information accumulation (i.e. decision making process) about the target picture. We also investigated whether the amount of attention deployed to words can affect the speed of perceptual decisions. We used a drift diffusion model (DDM) to analyze reaction time and two-alternative forced choice responses in an experiment with linguistic classification and verification tasks, where male and female faces were primed with gender labels 'man' and 'woman' (all participants were women). We showed that in the classification task, congruent prime-target pairs resulted in increased evidence accumulation in comparison to incongruent ones, following the well-established results from the literature (Voss, Rothermund, et al., 2013). In the verification task, female-related words resulted in an increased drift rate regardless of the gender of the target face, which suggests that the classification and the verification tasks require different performance strategies to accumulate decision about a visually presented category. Neural data for the classification task did not show any evidence in support of our hypothesis that the engagement of congruent representations can reflect an advantage in decision-making speed about the target picture. In the verification task, we found that the drift rate of evidence accumulation during incongruent versus congruent prime-target pairs with male faces correlated with activity in the precuneal cortex at the whole-brain level, which could potentially reflect the performance strategy during the task. This study offers a new perspective to the subject of language-perception interactions using a combination of model-based and neuroimaging approaches.

Introduction

We often think of the fundamental senses as being fixed. However, modern neuroscience tells us that the brain actively constructs experience based on prior beliefs and expectations. Language has been shown to exhibit top-down effects on visual perception (Carroll & Young, 2005; Gilbert et al., 2006; Landau et al., 2010; Winawer et al., 2007), and visually-driven decisions can be affected by various top-down factors such as expectations, attention, and previous beliefs (de Lange, Heilbron, & Kok, 2018; Gilbert & Sigman, 2007; Jehee, Brady, & Tong, 2011). For example, verbs congruently describing the direction of upcoming random-dot motion kinematograms (e.g., the word 'descend' followed by the dots moving downwards) resulted in faster and more accurate judgments about the presence or absence of the motion itself (Meteyard et al., 2007). In spite of the fact that the effect of language on perceptual decision making is well documented (see for review Raftopoulos & Lupyan, 2018), the locus of this effect is still a matter of a debate. While some studies have shown that language can modulate perceptual representations (such as fusiform face area known to reflect face processing, Amado et al., 2018; Puri, Wojciulik, & Ranganath, 2009), others have suggested that top-down language effects involve processes of lexico-semantic memory and/or decision making (Francken et al., 2015; Tan et al., 2008). In short, it remains unclear how and at which processing stage language cues (i.e., words briefly presented before the visual input) affect perceptual decision making.

Computational models of decision making help to elucidate the mechanisms underlying the effect of language on perception. Generally, perceptual decision making can be formalized as a process of accumulation of noisy sensory information until a decision threshold is reached and the motor command is executed (Gold & Shadlen, 2007; Mazurek, Roitman, Ditterich, & Shadlen, 2003; Smith & Ratcliff, 2004). Within this framework, the information introduced by language cues could affect a) the state of the decision variable before the onset of the evidence accumulation (i.e., starting point, which can be higher for female faces if one has a preference of judging faces as more female- than male-like); b) the rate of evidence accumulation (faster with the cue that is congruent with the visual stimulus, for example); or c) the amount of information that is required to make a decision (i.e., a lower decision threshold in the case when the cue is congruent with the visual stimulus, for example).

Neuroimaging studies on primates have shown that regions such as the lateral intraparietal area (LIP) (Shadlen & Newsome, 2001), superior colliculus (Ratcliff, Cherian, & Segraves, 2003) and frontal eye fields (Hanes & Schall, 1996) exhibit time-series activity similar to predictions of evidence accumulation models (Ratcliff

& McKoon, 2008a; Usher & McClelland, 2001). In human studies, it has been shown that frontal regions can reflect the process of evidence accumulation. For example, in order to form categorical decisions in multisensory environments, frontal regions such as the inferior frontal sulcus (IFS) have to accumulate audio-visual evidence from the auditory and visual sources (Noppeney, Ostwald, & Werner, 2010). However it has been also shown that other than frontal regions, temporal regions, for example, can reflect the rate of evidence accumulation. Recently, it has been shown that blood oxygen level dependent (BOLD) activity in face/house selective regions in inferior temporal cortex (IT, Haxby et al., 2001) was found to increase at a rate proportional to the decision time in favor of the preferred stimulus in the context of a face/house discrimination task (Tremel & Wheeler, 2015). Recent work on the identification of recurrent structures has shown that lesions to posterior middle temporal gyrus (pMTG) slow down information accumulation in the visual domain (Martins et al., 2019). The authors interpreted this finding in terms of available memory representations, which guide perception in a top-down manner. In sum, depending on the task, various brain regions may accumulate evidence that drives choice during decision making (Mulder et al., 2014). To our knowledge, it has not been shown which perceptual decision-making mechanisms can be affected by language, and whether any brain regions can exhibit correlations with discovered decision-making patterns evoked by language, if any.

In this study, we investigated how decision-making processes might affect the underlying language-vision interactions using an evidence accumulation framework in combination with neuroimaging methods (Ratcliff & McKoon, 2008a; Ratcliff, Smith, Brown, & McKoon, 2016). In light of previous studies highlighting the critical role of the speed of evidence accumulation in semantic decision tasks (Gomez et al., 2013; Voss, Rothermund, et al., 2013), we focused on investigating the link between the speed of evidence accumulation and the neural substrate underlying semantic decision making. We implemented a priming paradigm, in which a prime word (see Wentura & Degner, 2010) was followed by a briefly flashed visual input (or target) with the instructions to make a decision based on the target image. Primes were the words 'man' and 'woman' and targets were faces morphed from male to female in equidistant steps. Primes could be either congruent (e.g., prime: 'man', target: male face) or incongruent (e.g., prime: 'woman', target: male face) with the target. Behaviorally, we expected that congruent in comparison to incongruent pairs would result in faster reaction times and more correct responses (priming effect). Additionally, we modulated the amount of attention placed on the linguistic prime by introducing two tasks – classification and verification – that required differential allocation of attention. Classification is essentially a priming task, in which a word (i.e., the prime, see Wentura & Degner, 2010) is followed by a briefly flashed visual stimulus (i.e., the target), and in this case, the prime is uninformative for identifying the target item. In contrast, the verification task required an explicit match between the prime and target. Particularly, it was necessary to extract the meaning from both the prime and target and to assess whether they match while holding both items in working memory. The crucial difference between the tasks is that the prime word in the verification task cannot be disregarded, but it is rather essential to attend to the word for task performance.

Using functional magnetic brain resonance (fMRI), we were able to investigate whether the area associated with face processing such as fusiform face area (FFA) can be modulated by language, specifically by the semantics of the prime word in the context of a face categorization task. We hypothesized that priming effects (congruent > incongruent) could be found in behavioral, computational and neuroimaging measures. We further explored whether activity in the FFA could be accounted for by evidence accumulation in the DDM framework. We therefore tested whether BOLD in any of the brain regions found for the congruent vs. incongruent contrasts could predict the speed of the drift rate of evidence accumulation during decision making. We also investigated the differences between the two tasks. In the verification task we expected larger priming effects (congruent > incongruent) than in the classification task due to the fact that more attention is allocated to the prime. We also expected a similar set of modelling parameters reflecting the congruency effects as in the classification task, perhaps with larger effects. We also expected similar activation in FFA reflecting the congruency effect in classification and verification tasks.

To our knowledge, this is the first time that a neuroimaging study has addressed the language-based priming effects through the lenses of computational modelling.

Methods

We tested whether the lexico-semantic content of primes (i.e. ‘man’ vs. ‘woman’) affects the process of evidence accumulation during face-gender decisions (i.e. male vs. female) as a function of task demands, i.e., the requirement of no attention to the prime (classification task) or constant attention to the prime (the verification task). Decision-making mechanisms were evaluated by entering choice reaction time data for these decisions into the DDM. Crucially, we tested whether the resulting DDM parameter(s) could be used to predict brain activation using fMRI (whole brain and region of interest approach) in face-selective regions during the face gender decision tasks.

Subjects

Perception of faces, and particularly of face gender, has proven to be influenced by the gender of subjects (for example, see own-gender bias, McKelvie, 1987; Wright & Sladden, 2003). To avoid the gender effect in the regions associated with face perception, we opted for conducting our study on female subjects only (for gender differences in face categorization see: Coutrot, Binetti, Harrison, Mareschal, & Johnston, 2016; Lewin & Herlitz, 2002). Twenty-seven female volunteers recruited via the Radboud Research Participation System participated in the fMRI experiment. The study was approved by the local ethics committee (CMO Arnhem-Nijmegen, Radboud University Medical Center) and conducted in accordance with their guidelines. All participants gave written informed consent before the experiment and received standard monetary compensation for their participation. The data from five subjects were excluded from the analysis: three subjects failed to finish the task and two subjects exhibited head motion that exceeded the maximum acceptance rate of 2 mm. The final sample consisted of 22 subjects (Dutch native speakers, right-handed, age-range: 21-32, mean age = 25.86, *SD* = 3.07). All participants reported no neurological disease and had normal or corrected-to-normal vision.

Stimuli

The study required a set of faces as stimuli that would be categorized as male, female and ambiguous in a graded fashion. A set of realistic 3D gender-morphed faces was created with FaceGen Modeller 3.5 (Singular Inversions). The technical details on the computation method used by the software are discussed in (Blanz & Vetter, 1999b). We created 81 western face identities, for which we gradually morphed the gender features from extreme female to extreme male in ten equal steps (Figure 1A). The face stimuli were presented frontally and cropped around the oval of the face. We controlled for luminance using the SHINE toolbox for MATLAB (Willenbockel et al., 2010). The perceptual boundary within the gender continuum of faces was established in a separate behavioral experiment (see below).

To obtain subjective perceptual ratings of face gender on our generated images, we conducted a stimulus evaluation experiment in an independent sample. A group of 47 volunteers (24 males, native Dutch speakers, age range 20 - 32 years, mean age = 23.83, *SD* = 2.88) participated. Subjects performed a 2AFC discrimination task on the gender (male vs. female) of each of the faces from the set. Each trial in the evaluation experiment started with a fixation cross that stayed on the screen for 1500 ms, after which the target was visually presented for 500 ms. Participants were able to respond for an additional 2000 ms after the picture was removed from the screen. There were 810 trials in total (each identity was presented 10 times), which summed up to ~54 min of experimental duration. Participants had 2 optional

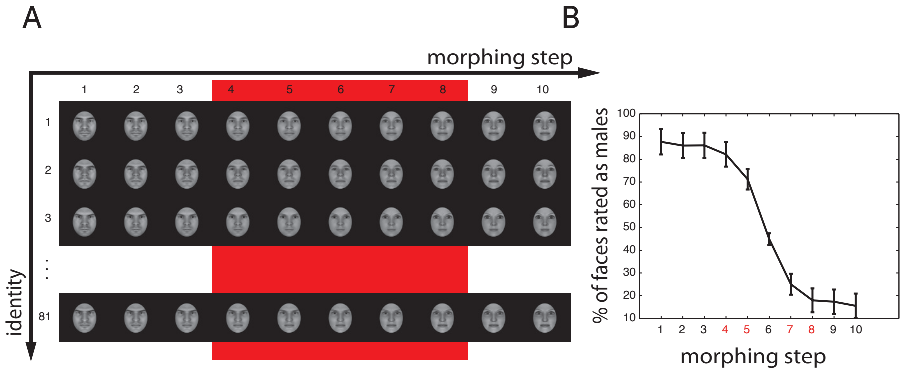


Figure 1 (A) Stimuli set with 81 identities each of which was morphed in ten steps between the male and female spectrum. The red square indicates the stimuli selected for the experiment, corresponding to faces of morphing steps 4-8 ('very male'(4), 'less male'(5), 'ambiguous'(6), 'less female'(7), and 'very female'(8), respectively). **(B)** Proportion of faces evaluated as males as a function of morphing step. The red numbers are morphing steps included in the experiment.

breaks during the experiment to avoid fatigue. Based on the responses, we identified the morphing step of faces that was perceived ambiguously. We included five morphing steps in the main experiment: two steps from the male/female continuum adjacent to the ambiguous point and one for ambiguous faces (Figure 1B). The percentage of faces evaluated as males for ambiguous faces (step 6 in Figure 1) was 47.87% (*SD* = 16.45), for the selected extreme-male faces (step 4) 84.21% (*SD* = 27.19), for the selected extreme-female faces (step 8) 16.80% (*SD* = 25.57).

Face-gender classification and verification tasks

Two different tasks (2AFC) were administered to each participant: a classification and a verification task. We set out to test whether ignoring the prime (classification) vs. keeping track of the prime at all times (verification) led to differential results on visual perception.

The classification and the verification tasks had an identical trial structure, and were both evaluated in terms of prime-target congruency (i.e., a prime word followed by a target image), but the instructions and response options differed between the tasks (see Figure 2 A, B in Results). For both the classification and the verification tasks (event-related design), each trial started with a prime (250 ms) that was either the gender-related word 'man' or 'woman' (Dutch: 'man' or 'vrouw'); after the prime word, there was a delay (fixation cross, 300 ms) followed by the target image of a face (500 ms) and the inter-trial interval (5-7 s, jittered).

In the classification task, participants were instructed to decide whether each face (target) was female or male ('man'/'vrouw' response); they were also encouraged to read but ignore the prime words as non-informative for their decision about the target image. In the verification task, participants were asked to perform a match between prime and target: respond 'yes' if a word and subsequent picture corresponded in gender, and 'no' otherwise. Participants were instructed to attend to the prime word in the verification task since it was essential for further matching with the target image.

In both the classification and the verification tasks, there were ten conditions determined by five levels of morphing from male to female faces (see Stimuli section for details) and two prime words ('man' or 'woman'). A congruent trial was defined as when both the prime word and target image had the same gender (e.g., 'man' followed by a *male* face). It is important to highlight that congruency was determined by the ratings in the first experiment, i.e., "very/less male/female" faces were evaluated as such in the first experiment. In contrast, an incongruent trial was defined as when the prime word and target image had different genders (e.g., 'man' followed by a *very female* face). Note that ambiguous faces were not considered as either congruent or incongruent. Within each block (and for each of the two tasks), the ten conditions were shown five times each, yielding 20 congruent and 20 incongruent trials per block (ten trials with ambiguous faces).

To ensure that participants read the prime word (this was particularly relevant for the classification task since gender categorization could be done without reading the word), we introduced 'catch' words in the classification task and instructed participants to skip trials containing any words other than 'man'/'vrouw'. Thirty-eight catch trials were introduced within a span of 4-6 trials in the classification task only, which summed up to ~10 additional trials per each of the classification blocks (~60 trials per each classification block).

Procedure

Participants performed two different two-alternative forced-choice (2AFC) face-gender decision tasks (i.e. classification and verification) while the BOLD signal was measured using fMRI in a single session (~1.5 hours). Participants stayed in the scanner for the duration of the session. The experiment was divided into 12 runs (6 runs for the classification and 6 runs for the verification task). The duration of each run was ~ 7 min. The 10 conditions within each run were presented randomly. Each run started with a brief instruction describing either the classification or the verification task. The order of the runs was randomized across participants such that the tasks did not repeat more than twice in a row. In total for each of the two tasks, 120 trials of both the congruent and incongruent conditions were administered per participant (60 trials of ambiguous faces).

We used the Presentation software (version 17.1, www.neurobs.com) in order to display the stimuli on a PC screen (Dell T5810, resolution: 1024x768, refresh rate: 60 Hz), which were further projected on the back of a transparent diffuse screen (DAP diffuus KBBA High Contrast, projection size: 454x340 mm, visual view angle H: 33 degree, V: 25 degree) via an LCD projector (EIKI LC XL100, resolution 1024x768, 60 Hz refresh rate) located outside the shielded MRI room. Participants replied with a response button (8 Button Bimanual Curved Lines HHSC-2x4-C) by pressing with the middle or index finger of the right hand. The buttons were counterbalanced across subjects.

At the end of the experimental runs (classification and verification tasks), participants performed one run of a face-selective localizer task in order to isolate voxels responsive to faces (~15 min.). In addition, we collected a T1-weighted anatomical image that lasted for approximately 5 min. We collected the T1-weighted image either before the main task or after the localizer task had been completed.

Behavioral analyses

We investigated effects of the task and the prime ('man'/'woman') on the categorization of the face gender target picture. We performed a repeated measures analysis of variance (rm-ANOVA) with within-participant factors: 'task' (levels: classification, verification), 'congruency' (levels: congruent, incongruent) and 'target' (levels: male, female). We collapsed across the morphing steps for each gender (i.e., male: *very male* and *less male* faces; female: *very female* and *less female* faces). We did not include the ambiguous faces in the analysis. Missed trials were excluded from the analysis. Planned comparisons included the following effects: congruent versus incongruent (separately per task). In addition, as an exploratory part of the analysis, we constructed gender-specific congruency contrasts (e.g., congruent vs. incongruent female only). We further performed the rm-ANOVAs on the dependent variables of reaction time (RT) and percentage of correct responses separately for the classification and the verification tasks. The analyses were performed using JASP (JASP Team, 2018).

Modelling analyses: Hierarchical Drift Diffusion Model (HDDM)

As mentioned in the introduction, our main goal was to investigate the link between the speed of evidence accumulation (i.e., drift rate) and the neural activity underlying the priming effects in the classification and verification tasks. In order to estimate the drift rate, we had to quantify computationally the processing components underlying gender judgments in the verification and classification tasks. To do so, we analyzed choice reaction-time data with the hierarchical drift-diffusion model. The analysis was implemented in the Python toolbox HDDM 0.6.0 (Wiecki et al. 2013). We used a hierarchical version of DDM (HDDM) where model parameters are estimated in a Bayesian framework. One of the main advantages of the hierarchical Bayesian framework is that the simultaneous estimation of the model parameters at both the subject and group levels enhances statistical power (fewer trials are required to recover the parameters and the estimates are less susceptible to outliers (Wiecki et al., 2013)), making it an appropriate analytic approach for the present study. Models with different combinations of free parameters were fit to the data via Markov Chain Monte Carlo (MCMC) fitting routines.

In the classification task, correctly categorized faces (male or female) were coded as 'correct' responses and errors as 'incorrect'. In the verification task, correct matches (e.g. prime: 'woman', target: 'woman', response: 'yes') and correct rejections (e.g. prime: 'woman', target: 'man', response: 'no') were coded as 'correct' responses; similarly incorrect matches (e.g. prime: 'woman', target: 'woman', response: 'no') and incorrect rejections (e.g. prime: 'woman', target: 'man', response: 'yes') were coded as 'incorrect' responses. We excluded from the analysis missed trials. For both decision tasks, we did not include the ambiguous faces in the analysis.

For each model, we evaluated the rate of convergence of the numerical fitting routines and then the ability of the model to capture the observed RT distributions. Models that failed to reach convergence or failed to capture the observed RT distributions were excluded from further analyses. Finally the models were compared against each other by computing the relative Deviance Information Criterion (DIC), which is a measure of the goodness of the model fit to the data that penalizes for the complexity of the model (Schwarz, 1978). We used a rm-ANOVA with congruency and target (see Behavioral analyses section) as within-subject factors to analyze parameter estimates of the best fitting model. The analyses were carried out using JASP (JASP Team, 2018).

fMRI acquisition, preprocessing & statistical analysis

Functional images were acquired using a 3T Skyra MRI system (Siemens Magnetom), T2* weighted echo-planar images (gradient-echo, repetition-time TR = 1760 ms, echo-time TE = 32 ms, echo spacing = 0.7 ms, 1626 hz / Px bandwidth, generalized auto-calibrating partially parallel acquisition (GRAPPA), acceleration factor 3, 32-channel brain receiver coil). In total, 78 axial slices were acquired (slice timing: interleaved; 2.0 mm thickness, 2.0 × 2.0 mm in plane resolution, 212 mm field of view (FOV) whole brain, anterior-to-posterior phase-encoding direction). In addition, we acquired T1-weighted high-resolution images (voxel size 1 × 1 × 1 mm) using a gradient echo sequence (TR = 2.3 sec, echo time = 3.03 ms, flip angle 8°, 256 mm FOV).

The data preprocessing and statistical analysis (the same preprocessing for classification, verification and localizer tasks) were performed using the FMRIB Software Library version 6.0.0 (FSL; https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/; Jenkinson, Beckmann, Behrens, Woolrich, & Smith, 2012; Smith et al., 2004; Woolrich et al., 2009). We carried out preprocessing steps using FEAT, version 6.00 (Woolrich, Behrens, Beckmann, Jenkinson, & Smith, 2004; Woolrich, Behrens, & Smith, 2004; Woolrich, Ripley, Brady, & Smith, 2001). For each run in the preprocessing step, we excluded the first 12 volumes needed for stabilization of the scanner's magnetic field. We applied grand-mean intensity normalization of the entire 4D dataset by a single multiplicative factor and high-pass temporal filtering (Gaussian-weighted least-squares straight line fitting, with sigma = 60.0s). We carried out motion correction using MCFLIRT (Jenkinson, Bannister, Brady, & Smith, 2002), and spatial smoothing (Gaussian kernel of FWHM = 6 mm). Functional images were brain-extracted using BET (Brain Extraction Tool; Smith, 2002).

For the first-level analysis of the classification and the verification tasks, we convolved the time-course of each trial-type regressor with the double-gamma hemodynamic response function. Voxel-wise time-series analysis was executed using FILM (FMRIB's Improved Linear Model; Woolrich et al., 2001) with autocorrelation correction. We registered each of the first-level contrasts to the high-resolution T1-weighted image of the corresponding participant using boundary based registration (BBR) as implemented in FLIRT (FMRIB's Linear Image Registration Tool; Jenkinson & Smith, 2001; Jenkinson et al., 2002). The EPI images were registered (via the high-resolution images) to the MNI-152 template with 12 degrees of freedom, applying a normal linear search. For each run at the first level, we constructed a general linear model (GLM) with ten regressors (five conditions of the face target and two conditions of the prime word). Contrasts of interest were the following (positive and negative directions): congruent vs. incongruent (male and female genders combined), male vs. female target, and the interaction between congruency and target (F-test). As an exploratory part of the analysis, we constructed four gender-specific congruency contrasts: congruent female vs. incongruent female and congruent male vs. incongruent male. Six rigid-body motion parameters were included as nuisance regressors.

At the second level, the resulting contrast images from the first-level analysis were averaged across runs for each participant for the classification and the verification tasks separately (fixed effects modeling, FEAT). The third level (group), was carried out using FLAME (FMRIB's Local Analysis of Mixed Effects) with Z-statistic images thresholded at $Z > 3.1$ and a cluster-corrected threshold of $p < .05$ to control for the family-wise error rate.

Finally, in a separate group-level analysis, we tested for a correlation (positive and negative directions) between congruency effects (congruent > incongruent) in the (demeaned) HDDM parameters and BOLD contrast of parameter estimates for the classification and the verification tasks separately. Here, we focused on the drift rate, and performed further exploratory analyses on the threshold and starting point. We also tested a correlation (positive and negative directions) between significant gender specific effects and respective BOLD contrasts. The correlations were evaluated at the level of the whole brain data using FLAME (FMRIB's Local Analysis of Mixed Effects) Stage 1. Statistic images (Z-statistics) were thresholded at $Z > 3.1$ and a cluster-corrected threshold of $p < .05$ to control for the family-wise error (FWE) rate (Eklund, Nichols, & Knutsson, 2016).

Defining face-selective regions of interest

For a region of interest (ROI) analysis, we sought to isolate voxels responding selectively to faces. To do so, we made use of a functional localizer (for details see below). Although the exact neural substrate that reflects face gender perception is still an area of active investigation (Kaul, Rees, & Ishai, 2011; Wiese, Kloth, Güllmar, Reichenbach, & Schweinberger, 2012), it has been unambiguously shown that one region in particular, the FFA, largely contributes to the coding of gender (Contreras, Banaji, & Mitchell, 2013; Freeman, Rule, Adams Jr, & Ambady, 2010). Here, rather than investigating the neural representation of face gender, we focus on FFA as a brain area, which has proven to be a major contributor to face-gender coding.

The face localizer was a one-run task, which had a blocked design with four conditions: faces, houses, scrambled faces and scrambled houses (see Supp. Figure 3). Conditions were presented in a randomized order. Within each block, 20 images were shown for 1 s, preceded by a fixation cross for 1 s (40 s per block). Each block was repeated four times in a single run. There was a jittered 6000-12000 ms break after each block while scanning continued. In the localizer, a total of 80 unique images were shown to each participant. While passively viewing the stimuli during the localizer participants had to decide whether two stimuli repeated twice in a row and deliver a response if so.

We defined the FFA at the group level. All preprocessing stages were the same as for the main tasks. At the first level we defined four regressors (faces, houses, scrambled houses, scrambled faces) and modelled the contrast of interest as follows: faces > houses + scrambled houses. At the second (and group) level, the resulting subject-specific parameter estimates were evaluated (FLAME, $Z > 3.1$, $p < 0.05$ FWE). Face-selective voxels were initially restricted based on a combined anatomical mask of the Temporal Fusiform Cortex (posterior division) and the Temporal Occipital Fusiform (see Figure 4) using the Harvard-Oxford Cortical Structural Atlas (as part of FSL). Mean activation for each contrast of interest within the FFA was computed using a weighted-means approach, based on each voxel's group-level activation (thresholded posthoc at $Z > 4.0$ in order to increase the specificity of the cluster selection). We evaluated the mean Z-statistics in a rm-ANOVA with factors: congruency (levels: congruent, incongruent), target (levels: male, female) and task (levels: classification, verification). We also performed linear regression to investigate whether BOLD activation in FFA can predict the drift rate parameter modelled in the DDM framework. The analysis was carried out using JASP (JASP Team, 2018).

Results

Behavioral analyses

First, we tested using RT and correct responses whether face categorization differs in the context of classification and verification tasks (see Table 1 and Figure 2). In RTs, we found an interaction between task and target and a three-way interaction between task, congruency and target (see Table 1). In correct responses, we found a main effect of task and an interaction between task and congruency (see Table 1). We further explored the effect of prime separately per classification and verification tasks.

In the classification task (Fig. 2A), we found that congruent prime-target pairs resulted in faster RTs as compared with incongruent prime-target pairs: $t(21) = -4.08$, $p < .001$. We found that trials with congruent female targets resulted in more accurate categorization judgments: $t(21) = 3.13$, $p = .003$. We did not find a significant difference between congruent vs. incongruent trials with male targets: $t(21) = 1.25$, $p = .112$.

In the verification task (Figure 2 B), we found slower RTs for incongruent as compared with congruent prime target pairs for female targets: $t(21) = 4.39$, $p < .001$. We did not find any difference between incongruent as compared with congruent prime-target pairs with male targets: $t(21) = .27$, $p = .391$. In the accuracy data, we found that

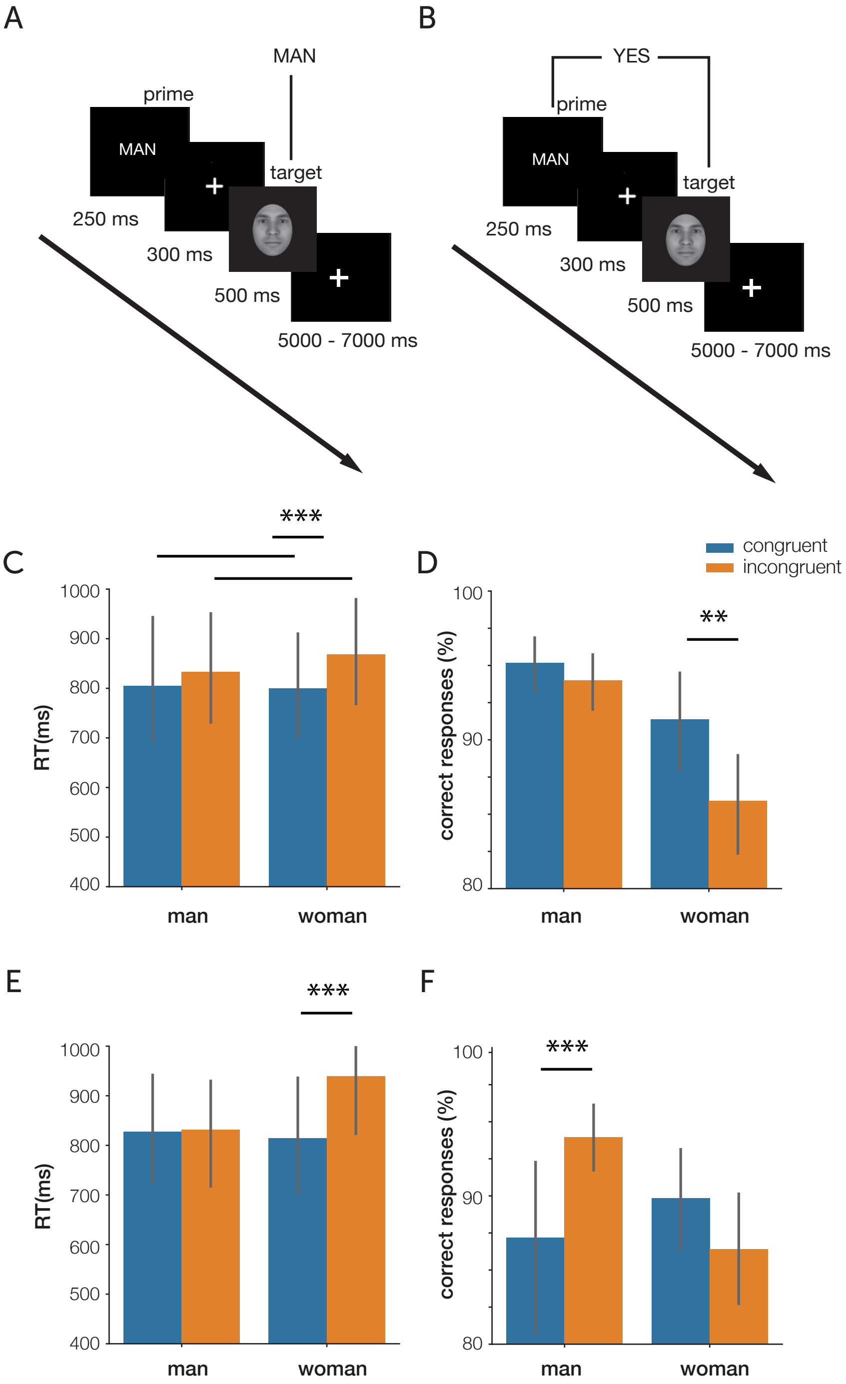

A
MAN
prime
MAN
target
250 ms
300 ms
500 ms
5000 - 7000 ms
B
YES
prime
MAN
target
250 ms
300 ms
500 ms
5000 - 7000 ms
C
***
1000
900
800
700
600
500
400
RT(ms)
man
woman
D
congruent
incongruent
**
100
90
80
correct responses (%)
man
woman
E
***
1000
900
800
700
600
500
400
RT(ms)
man
woman
F
***
100
90
80
correct responses (%)
man
woman

Figure 2: Classification and verification tasks. Both tasks consisted of the same trial structure (prime followed by target, two-alternative forced-choice), while the instructions and response options differed between the tasks. These two tasks were designed to test whether engagement of the prime in the task had differential effect on gender categorization. **(A)** The classification task in which participants discriminated the face gender of the target (options: male or female). They were instructed to read the prime word but were told it was non-informative for discriminating the target image. (**B**) The verification task, in which participants were instructed to decide whether the gender described in the prime matched the face gender of the target image (options: yes or no). (**C, D**) Performance results reflecting face (target) classification as a function of prime-target congruency during the classification task and (**E, F**) the verification task. (**C, E**) Error bars, standard error of the mean. ** $p < 0.01$, *** $p < 0.001$

participants were more accurate for incongruent as compared with congruent trials for male targets: $t(21) = 3.85$, $p < .001$. We did not find a significant difference between congruent vs. incongruent trials for female targets: $t(21) = 1.68$, $p = .053$.

In sum, in the classification task, participants were faster for congruent as compared with incongruent faces and more accurate for congruent prime-target pairs with female faces, but not with male faces. In the verification task, participants were

Table 1 Repeated measures ANOVA (2-way) results for RTs and percent correct.

RT	*F*(1,21)	*p*	*η*²
Task	*2.27*	*.146*	*.030*
Congruency	*20.55*	*< .001****	*.131*
Target	*14.59*	*<.001****	*.040*
Task * Congruency	*1.194*	*.287*	*.003*
Task * Target	*6.526*	*.018**	*.011*
Congruency * Target	*14.93*	*<.001****	*.066*
Task * Congruency * Target	*4.86*	*.039**	*.016*
Percentage Correct			
Task	*7.075*	*.015**	*.027*
Congruency	*.537*	*.472*	*.004*
Target	*15.854*	*<.001****	*.092*
Task * Congruency	*6.613*	*.018**	*.032*
Task * Target	*3.068*	*.094*	*.016*
Congruency * Target	*12.593*	*.002***	*.069.*
Task * Congruency * Target	*3.064*	*.095*	*.012*

Factors tested were task (classification, verification), congruency (levels: congruent, incongruent) and target (levels: male, female). * $p < .05$; ** $p < .01$; *** $p < .001$. CL – classification task, VR – verification task.

slower when they saw incongruent prime-target faces with female targets in comparison to congruent pairs with female targets (no effect for male targets). We also found in the verification task that participants were more accurate for incongruent pairs with male targets as compared with congruent pairs with male targets (no effect for female targets).

Modelling analyses

We performed drift diffusion analysis using RTs and choice responses separately for the classification and the verification tasks. Our research hypothesis was that better subject performance in congruent vs. incongruent conditions of semantic decision tasks could be explained by the modulations of the drift rate. Therefore we focused on the drift rate out of all modelling parameters (see Figure 3). Exploratory analyses concerning decision threshold and starting point are presented in the Supp. Fig. 1 and Supp. Table 1.

Model Convergence and Model Fit

For all of the analyses reported, the MCMC (Gelman & Rubin, 1992) fitting routines were run for 20.000 iterations with a burn-in period of 10.000 iterations and a thinning of 5. Model convergence was assessed by examination of the posterior samples and of the R-hat statistic, which is a measure of convergence among multiple MCMC chains (three for the present study). Posterior density estimates, which are stable over multiple samples, indicate that the fitting routines have converged to a fixed estimate. Similarly, an R-hat statistic below 1.1 indicates then that chains with different starting values have converged to the same posterior estimate. Successful convergence was further confirmed by a MCMC error for all of the parameters smaller than 0.01. We further performed a comparison between observed and recovered RT distributions produced by the model (see Supp. Figure 2 for an example of model fit for one condition). After assessing convergence, we carried out a quantitative comparison of alternative models by computing the associated DIC score for each model. DIC is a measure of the goodness of fit of the model to the data that is penalized for the complexity of the model. Therefore, a model with a lower DIC score is to be preferred over an alternative model with a higher DIC as the most parsimonious explanation of the data. Models that did not reach convergence were discarded and not included in the DIC comparisons.

The model that best described the data (i.e., the model with the lowest DIC score, see Supp. Table 2 for details) for the verification task was the model with the following parameters estimated per subject: drift rate (*v*), non-decision time (*Ter*), decision threshold (*A*) and starting point (*z*) free to vary over *prime* and *target* conditions. In other words, there was no constraint for these parameters to be

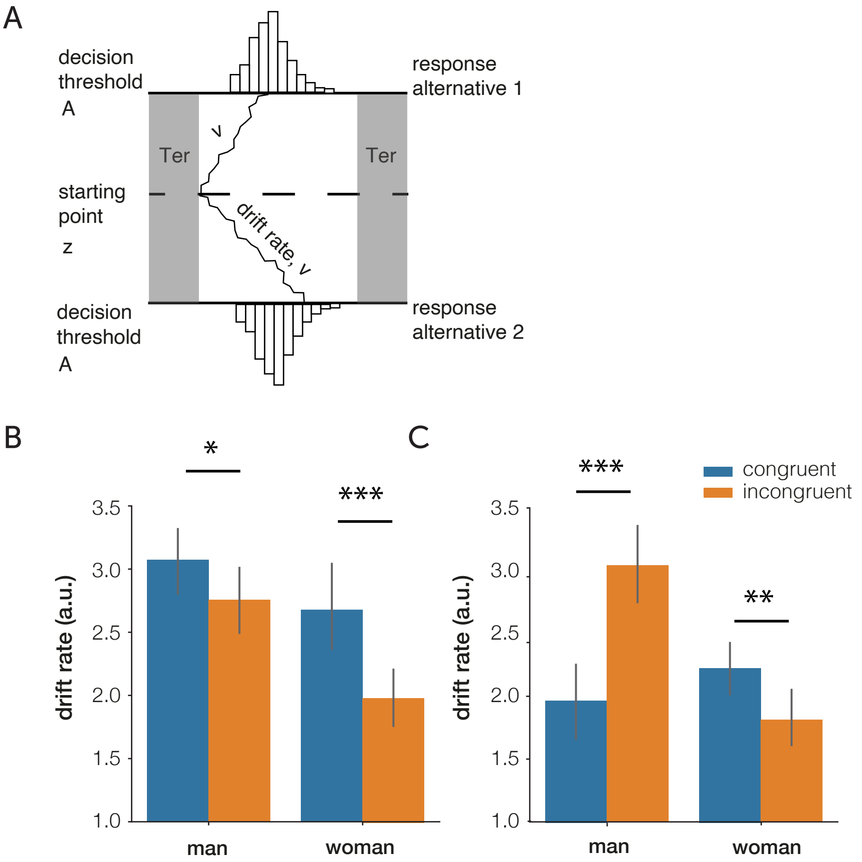


Figure 3: Schematic representation of the drift diffusion model. Once the stimulus is encoded, the process of evidence accumulation begins. The evidence is accumulated towards one of the decision boundaries. After the decision threshold is reached, the response is executed. **(A)** The Drift Diffusion model with the four parameters: drift rate (*v*), decision threshold (*A*), starting point (*z*), non-decision component (*Ter*). Drift rate estimates for classification **(B)** and verification **(C)** tasks. Error bars, standard error of the mean. ** $p < 0.01$, *** $p < 0.001$

constant for some specific conditions (*prime* or *target*).Conventionally, a DIC difference of more than 10 indicates that the evidence in favor of the model-winner is substantial (Burnham, Anderson, & Burnham, 2002). Because the difference between the winning model (model 21, DIC 710.319) and the second-best model (model 19, DIC 740.89) exceeds 10, we consider this evidence sufficient for selecting model 21 as the most parsimonious account of the data and therefore further analyses focus on this model.

For the classification task, the best model was the model with the following parameters estimated per subject: drift rate (*v*), non-decision time (*Ter*), and decision threshold (*A*) free to vary over *prime* and *target* conditions. The difference between the first three top models was minimal (model 21: -210.14, model 0: -210.26, model 20: -208.32). Since adding additional complexity (models 20 and 21 had the z parameter in addition to v, A, Ter) did not result in substantial advantages in model fit (DIC did not exceed 10 scores between the models), we selected model 0 (with v, A, Ter parameters) as the most parsimonious account of the data and further focused on analyses with this model.

Model parameter analysis

We performed a rm ANOVA analysis with congruency and target face as factors. Results are summarized in Figure 3 and Table 2.

Table 2 Repeated measures ANOVA (2-way) results for drift rate parameter. Factors tested were congruency (levels: congruent, incongruent) and target (levels: male, female). * $p < .05$; ** $p < .01$; *** $p < .001$. CL – classification task, VR – verification task.

Drift rate	$F(1,21)$	p	η^2
Task	*17.93*	*<.001****	*.065*
Congruency	*.664*	*.424*	*.004*
Target	*30.982*	*<.001****	*.143*
Task * Congruency	*30.681*	*<.001****	*.088*
Task * Target	*.385*	*.542*	*.001*
Congruency * Target	*30.927*	*<.001****	*.109*
Task * Congruency * Target	*13.194*	*.002***	*.038*

Drift rate (V)

We found a main effect of task, an interaction between task and congruency and a three-way interaction between task, congruency and target (see Table 2). We further present the effect of primes for classification and verification results separately. For the classification task (Figure 3 C), for both female and male faces, congruent primes led to increased drift rate in comparison to incongruent pairs (female targets: $t\ (21) = 3.82$, $p < .001$; male targets: $t\ (21) = 2.63$, $p = .016$).

For the verification task (Figure 3 D), we found increased drift rate for incongruent vs. congruent pairs with male targets *t* (21) = 4.60, *p* < .001. For the female targets, the drift rate was increased for congruent vs. incongruent targets *t* (21) = 3.21, *p* = .002.

Results of the exploratory analysis for decision threshold, non-decisional component and starting point are presented in the Supplementary Figure 1, Supp. Table 1.

To sum up, drift rate in the classification task followed the priming pattern of results – increased drift rate for congruent in comparison to incongruent prime target pairs. In the verification task, we unexpectedly observed another pattern: increased drift rate for incongruent vs. congruent prime-target pairs for male faces and increased drift rate for congruent vs. incongruent prime-target pairs with female faces in comparison. This might suggest that participants strategically performed the verification task differently from the classification task, which can be explained by paying more attention to the prime "woman" vs. "man", which resulted in higher drift rate. It is a prominent finding that attention increases the speed of sampling information, which can be captured via increased drift rate (Armel, Beaumel, & Rangel, 2008; Krajbich, Armel, & Rangel, 2010; Lim, O'Doherty, & Rangel, 2011). Particularly, it has been suggested that attention accelerates the entry of the stimuli into short term memory, which is reflected in the drift rate (Smith, Ratcliff, & Wolfgang, 2004). In this study, in the verification task, allocation of attention to a word "man" / "woman" might create its representation in short term memory, and information from the images can be matched against it. This process is potentially reflected in the observed pattern of the drift rate.

Neuroimaging analyses

We tested whether gender-related word primes ('man' vs. 'woman') modulated brain activation while participants categorized faces as either male or female. BOLD-signal changes were evaluated at the whole-brain and ROI levels. A functional localizer for face-selective voxels was utilized in the current sample for defining the FFA (see Methods).

Localizing face-selective regions of interest

The contrast faces > houses + scrambled houses revealed a network of significantly activated brain regions (see Figure 4 and Table 3). We expected the face localizer to yield occipito-temporal brain regions that were previously shown to be related to face gender processing: LOC and TFC (Dricot, Sorger, Schiltz, Goebel, & Rossion, 2008; Hermann, Bankó, Gál, & Vidnyánszky, 2015; Kanwisher & Yovel, 2006).

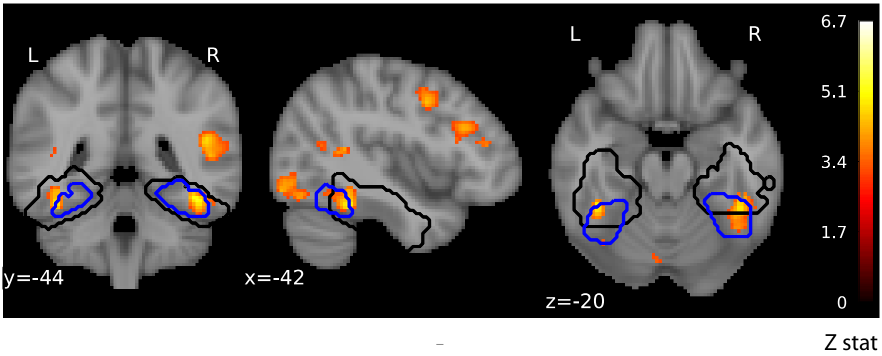


Figure 4: Face Localizer. Faces > houses + scrambled houses contrast, Z-statistics after cluster-correction. The outlines are anatomical masks from the Harvard-Oxford Cortical Structural Atlas that were used to restrict the selection of voxels from the localizer. In black: Temporal Fusiform Cortex, posterior division; in blue: Temporal Occipital Fusiform.

Table 3 Significant clusters of BOLD activation for the faces > houses + scrambled houses contrast in the localizer task. Results are cluster-corrected using FWE rate of 0.05. For each cluster, we present the Harvard-Oxford Cortical Structural Atlas coordinates (x, y, z) of the local maxima (MNI-152 2-mm space), including the cluster size (k), and the maximum Z-statistic at those coordinates.

Brain region	x	y	z	Z-stat	k
Frontal Pole	34	42	20	4.48	2287
Lateral Occipital Cortex	44	-84	-4	6.75	1568
Middle frontal Gyrus	-36	24	24	5.35	539
Cerebellum (Left Crus II)	-10	-78	-38	4.66	430
Temporal Fusiform Cortex	-42	-44	-20	5.17	406
Precentral Gyrus	-42	2	40	4.38	223
Precuneus Cortex	2	-58	34	4.1	169
Paracingulate Gyrus	-4	20	46	3.62	141
Right Amygdala	20	-4	-14	5.56	136
Angular Gyrus	-44	-58	16	4.22	133

Prime-target interactions in the classification and verification tasks

To explore prime-specific modulations in the BOLD signal, we performed whole brain rm-ANOVAs (congruency by target) separately per task. We did not find any effects using a whole brain analysis (see Supp. Table 3).

Since the verification task exhibited quite different patterns of results for female and male targets in the behavioral and modeling data, we performed an exploratory analysis for the verification task only, in which we investigated the following contrasts: congruent female > incongruent female; incongruent female > congruent female; congruent male > incongruent male; incongruent male > congruent male). However, we did not obtain significant results in any of the exploratory whole-brain analyses (Supp. Table 4).

The ROI-based analysis, within the FFA, based on a rm-ANOVA with factors congruency, target and task did not reveal any main effects or interaction effects (see Supp. Table 5, 6).

The relationship between drift rate and prime-target congruency

We performed a linear regression to investigate whether drift rate can be predicted by the BOLD activation related to prime-target congruency within the FFA. For the classification task, we performed one regression, i.e. drift rate ~ FFA (congruent > incongruent) regardless of the target. In the verification task, we performed two linear regressions separately for male and female targets, i.e. drift rate ~ FFA (congruent female target > incongruent female target) and drift rate ~ FFA (incongruent male target > congruent male target). We did not find that congruency effects within the FFA can predict drift rate in either the classification task ($B = -.01$, $t = -.32$, $p = .74$, $R^2 = .00$), or in the verification task (male targets, incongruent male target > congruent male target: $B = -.10$, $t = -1.19$, $p = .24$, $R^2 = .06$; female targets, congruent female target > incongruent female target: $B = .06$, $t = 1.33$, $p = 0.19$, $R^2 = .08$).

To explore if there are other regions that could correlate with drift rate, we performed a whole-brain correlation between BOLD signal and drift rate for the congruent > incongruent contrast in the classification task. In the verification task, we performed a whole-brain correlation analysis between BOLD signal and drift rate for incongruent male target > congruent male target and for congruent female target > incongruent female target. We obtained a positive correlation between the drift rate and BOLD activation for incongruent male > congruent male faces, which was localized to the left precuneus (see Figure 5; max. Z-statistic = 3.84 [-12x, -56y, 8z], 135 voxels). No other correlations were obtained (see Supp. Table 6).

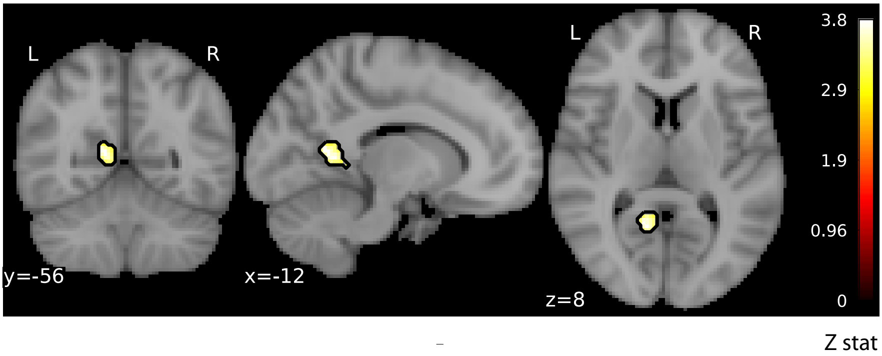


Figure 5: Significant BOLD activation for incongruent male faces > congruent male faces that correlates with the drift rate of evidence accumulation in the verification task. Color bar denotes Z statistics.

Discussion

In this study, we used a combination of neuroimaging (fMRI) and computational modeling (DDM) to investigate mechanisms of language-vision interactions from a decision-making perspective. Particularly, we studied whether semantic-related BOLD effects (e.g., the effect of the word 'male' on perception of a male face) in sensory cortex can be accounted for by evidence accumulation mechanisms. Further, we studied whether the amount of attention allocated to language semantics modulates the sensory BOLD signal and whether it is associated with the speed of evidence accumulation. For that, we constructed the tasks in which subjects had to perform a semantic decision by either ignoring the context (i.e., the prime in the classification task) or keeping track of the context at all times (i.e., the prime in the verification task).

Behavioral data showed that in the classification task, subjects were faster and more accurate in congruent as opposed to incongruent word-face pairs. In the verification task, the RTs and accuracies varied depending on the target face. We found shorter RTs for congruent in comparison to incongruent female faces (no difference between congruent/incongruent pairs for male faces). In percentage of correct responses, participants were more accurate for incongruent male faces in comparison to congruent male faces (no difference in accuracy performance for female faces). DDM modelling showed that the classification and the verification tasks resulted in differential modulation of the drift rate. In the classification task, congruent prime-target pairs resulted in increased drift rate in comparison to incongruent ones (for both female and male targets). In the verification task, drift rate varied depending on the target face similarly as it was shown in behavioral results. In male targets, congruent prime-target pairs resulted in increased drift rate in comparison to incongruent ones. In female targets, congruent prime-target pairs resulted in increased drift rate in comparison to congruent ones. While we did not find any correlation between drift rate and BOLD activation in the classification task, we found that in the verification task the BOLD activation in left precuneal cortex was positively correlated with the drift rate for incongruent male faces > congruent male faces at the whole-brain level. We interpret these findings in light of relevant literature below.

Language-perception interactions via an evidence-accumulation lens

According to the theories of grounded cognition (Barsalou, 2008; Gallese & Lakoff, 2005; Kiefer & Pulvermüller, 2012), conceptual knowledge is mapped onto sensory-motor representations. A number of neuroimaging studies reported activation in sensory-motor cortices while people were performing semantic tasks related to sensory or/and motor concepts. For example, color naming (as well as reading color words or words with strong color associations) resulted in activation of perceptual representations related to color (Bannert & Bartels, 2013; Brouwer & Heeger, 2013). Similarly, processing action-related words resulted in activations in motor cortex (Hauk, Johnsrude, & Pulvermüller, 2004; Klepp, van Dijk, Niccolai, Schnitzler, & Biermann-Ruben, 2019 but see de Zubicaray, Arciuli, & McMahon, 2013). An alternative to the grounded cognition approach, evidence accumulation, suggests that sensory-motor cortices rather than representing perceptual/motor information per se, can reflect the speed within which the brain has to accumulate information until the decision threshold is reached (Gold & Shadlen, 2007; Smith & Ratcliff, 2004). Indeed it has been shown that content-specific evidence accumulation for the preferred response in simple object discrimination is related to BOLD activity in temporal and occipital brain regions (see review Mulder et al., 2014).

In our study, we had two tasks: classification and verification. Regarding the former, as we mentioned earlier, classification is essentially a priming task, where participants decide about the gender of the face preceded by the gender-related or non-related word. Usually primed vs. unprimed stimuli result in decreased neural activity in brain areas engaged in the stimuli processing (Copland et al., 2003; Henson, 2003; Rissman, Eliassen, & Blumstein, 2003; Rossell, Price, & Nobre, 2003). Mechanisms of priming, in spite of years of research, are yet a matter of discussion (see review Henson, 2003). Here, we expected to find decreased brain activity for congruent as compared with incongruent items, which would correlate with the speed of evidence accumulation (see motivation above). In the classification task we did not find brain regions in which BOLD activity correlated with evidence accumulation. This suggests that the priming effect (i.e., BOLD activity in FFA for congruent > incongruent) cannot be easily explained in terms of the evidence accumulation framework. However, it can be that our study did not have sufficient power to reveal effects of interest since we did not find any results using BOLD only. With a block design and increased number of trials per condition, there might be higher chances to detect a significant correlation between model parameters and BOLD.

In the verification task, where participants have to match prime and target, we found that the speed of evidence accumulation positively correlated with BOLD activity in the left precuneal cortex. All of the participants were women in the current study. It can be that women decide about the gender of male and female faces in different ways. Particularly, male faces preceded by female words in comparison to male faces preceded by male words may trigger self-referential thinking (Moody et al., 2015). In other words, when deciding if prime "man/woman" is a match/mismatch to the target male face, participants use the word as a reference to match the target to. This results in an increased rate of evidence accumulation for male targets with female primes since there might be more interference matching out-group face to in-group word in comparison to matching out-group face to out-group word (for perception of female gender by females, see McKelvie, 1987; Wright & Sladden, 2003; Ino, Nakai, Azuma, Kimura, & Fukuyama, 2010). Yet another plausible explanation can be that the drift rate-BOLD correlation in precuneal cortex can reflect buffered memory regarding the upcoming response (Oishi et al., 2005), which can be more demanding for matching the words to the out-group face. Regardless of the exact role of precuneus, the results of the verification task highlight the social component in language-perception decision making.

This study has a few limitations. The first limitation concerns statistical power of the study. A follow-up study would benefit from increasing the number of trials per condition by having only one task, and without the ambiguous condition. The second limitation concerns the suboptimal localizer design. In the current version of the experiment, the localizer targets representations of faces rather than representations of face gender. Including male and female faces in separate localizer blocks will help to address this issue. Third, including male subjects in the experiment would allow disentangling better effects related to subject's gender from effects related to general aspects of cognition. Further methodological improvements can be achieved as follows. For example, ROIs can be defined more precisely at the single-subject level in order to account for inter-subject variability in the BOLD signal. Furthermore, a promising analytical direction is to estimate both drift rate and BOLD at the single trial level, which would further allow performing trial-by-trial correlations. This might lead to more precise conclusions about the relationship between the speed of evidence accumulation and gender specific BOLD representations.

To sum up, words tap into different mechanisms of information accumulation depending on the task. In the classification task congruent words in comparison to incongruent ones resulted in increased evidence accumulation, following the pattern reported in previous priming experiments. In the verification task, female word irrespective of the target face resulted in increased rate of evidence accumulation, which suggests social aspects at play while executing the task. This study contributes to the body of the literature dedicated to language-perception interactions and semantic-based decision making in general.

4

Post-perceptual effects of words on ambiguous visual perception: an MEG study

Abstract

While the top-down effects of language on visual perception are well demonstrated (Carroll & Young, 2005; Meteyard et al., 2007; Mitterer, Horschig, Müsseler, & Majid, 2009), the level at which language influences visual decisions is still a matter of controversy. Linguistic effects on visual perception have been observed both at early stages of visual processing (Boutonnet & Lupyan, 2015) and at later stages of lexico-semantic memory processing (Francken et al., 2015). In the case of ambiguous visual stimuli evidence is still scarce. To the best of our knowledge, there is only one study that shows that effects of lexical semantic memory (introduced by words) influence early visual processing (Samaha et al., 2018). However a study of perceptual decision making that involves classification of degraded pictures showed that both early visual and late decision components can underlie performance (Philiastides & Sajda, 2006). The goal of this study was two-fold. First, we investigated whether words influence ambiguous visual perception by modulating early or later stages of processing, which might be related to visual perception or semantic memory. Second, we set out to test whether these early or late effects (if any) would map onto these different computational profiles of the decision making process. Specifically, we investigated the effect of language on ambiguous visual perception within the mathematical framework of sequential sampling models (Forstmann, Ratcliff, & Wagenmakers, 2016). In this theoretical framework, the process of making a decision among two choices is described as the accumulation of evidence towards a threshold or boundary. Once a boundary is reached, the associated response is produced (Ratcliff et al., 2016; Smith & Ratcliff, 2004). Previous work has shown that regions in occipital and temporal brain areas can exhibit a pattern of activity which is reminiscent of a process of evidence accumulation (see for review Mulder, van Maanen, & Forstmann, 2014). Therefore, in the present study we investigated whether any of the language related effects would reflect the speed of evidence accumulation about visual decisions. To do so we used an hybrid approach combining the excellent temporal resolution of magnetoencephalography (MEG), which allows for the separation of early (visual) and late (lexico-semantic) processing stages, with the computational framework of the drift diffusion model (DDM, (Ratcliff & McKoon, 2008a)). Participants in this study were presented with ambiguous pictures of faces and houses preceded by either face-related, house-related or neutral words. Immediately after, they were asked to judge whether the picture just presented was a face or a house. Results showed a post-perceptual effect in event-related fields (ERFs) about 300 ms after the picture onset (P300). Contrary to our expectations, in the current paradigm words influenced ambiguous visual perception by modulating post-perceptual

mechanisms related to attention or decision rather than early visual or lexico-semantic mechanisms. We did not find correlations between P300 amplitudes and model parameters, which suggests that the process of evidence accumulation cannot be easily associated with the P300. The implications of this finding and avenues for future research are discussed.

Introduction

Words improve our ability to perceive things in the visual world. For example, the brief presentation of the word "dog" speeds up the visual recognition of familiar animals in comparison to the nonverbal sound of a barking dog (Boutonnet & Lupyan, 2015). When the picture is ambiguous, such as a two-toned Mooney face (Mooney, 1957), verbal hints about the possible interpretation of the picture improve visual discrimination performance (Samaha et al., 2018). Indeed, a growing body of evidence suggests that words exert top-down effects not only on non-ambiguous visual input (see for review Anderson, Chiu, Huette, & Spivey, 2011; Lupyan, 2012) but also on ambiguous ones: words boost otherwise invisible images into awareness (Lupyan & Ward, 2013) and aid recognition of ambiguous images (Samaha et al., 2018). While there is some evidence that language taps directly in the processes associated with early visual perception (see evidence in favor of that (Boutonnet & Lupyan, 2015; Samaha et al., 2018)), others propose that language has an effect on memory or decision processes (Francken et al., 2015; Mitterer et al., 2009; Tan et al., 2008). It is still not entirely clear how and at which processing stage language modulates perception of ambiguous visual input.

In this study we investigated whether words influence the perception of ambiguous pictures by influencing early visual or late processes associated with lexico-semantic memory. To be able to disentangle these processing stages, we analyzed the time-course of language effects on visual categorization with magnetoencephalography (MEG) capitalizing on its high temporal resolution. We utilized two well-known neural signatures to address this: the M170 and N400m in the event-related fields. The M170 is characterized by a negative amplitude that peaks at around 170 ms after the onset of the stimulus over occipital channels (Liu, Harris, & Kanwisher, 2002) and it is known to be a marker of face categorization and recognition. By contrast, the N400m (the MEG counterpart of the N400 event-related component) is known to manifest from the onset of the stimulus up until 500 ms after the stimulus onset in the temporo-occipital areas (Kutas & Federmeier, 2011; Lau, Phillips, & Poeppel, 2008) and is thought to reflect the processing of semantic meaning more broadly.

In this study we used a priming paradigm in which a briefly flashed word (face-related, house-related or neutral) was followed by an ambiguous picture created by superimposing a face and a house picture. Our interest lied on testing whether the amplitude of the M170 and/or N400m would be modulated by the semantics of the prime words. In particular, we tested whether ambiguous pictures (perceived as face) preceded by face-related words were processed differently than when preceded by house-related (or neutral) words. As an exploratory part of the analysis, we also investigated whether pre-target oscillatory activity set by the lexico-semantic primes could bias the upcoming perception of the target picture. In light of recent discoveries showing that language prepares the brain for the processing of upcoming perceptual stimuli by increasing pre-stimulus oscillations in the alpha range (Mayer, Schwiedrzik, Wibral, Singer, & Melloni, 2016; Samaha et al., 2018), we expected increased pre-stimulus alpha activity for stimuli that were congruent with the response (ambiguous pictures perceived as faces with face-related primes and ambiguous pictures perceived as houses with house related primes) in comparison to incongruent stimuli (ambiguous pictures perceived as faces with house-related primes and ambiguous pictures perceived as houses with face related primes).

Finally, we set out to test which of the effects (if any) among N170 or N400m would be accounted for by the computations of the decision-making process. In the neuroscience literature, well known effects are often attributed to certain cognitive processes, such as N400 effects to semantic processing (Kutas & Federmeier, 2011). However, in recent years a growing body of literature has suggested that activity in both temporal and occipital brain regions during binary decision making tasks could be associated with decision making dynamics (see for review Mulder et al., 2014) rather than pure semantic processing. Computational models of decision making characterize the process of making a decision about visual input (face or house) as the stochastic accumulation of evidence towards a decision threshold (Gold & Shadlen, 2007; Mazurek et al., 2003; Smith & Ratcliff, 2004). Here we used the framework of sequential sampling models (Ratcliff et al., 2016; Smith & Ratcliff, 2004) to formally characterize and separate the different stages of processing underlying the forced-choice binary decisions (face or house). From this point of view, the advantage provided by language in visual categorization can translate into a) the state of the decision variable before the start of evidence accumulation; b) the rate of evidence accumulation (faster for congruent prime-target pairs in comparison to incongruent ones); c) the amount of evidence needed for evidence accumulation (lower threshold for congruent vs. incongruent prime-target pairs). We further tested whether the effects of linguistic primes on visual categorization (M170, N400m) can be accounted for by the process of evidence accumulation.

In conclusion, we used a hybrid approach comprising of a neuroimaging technique (MEG) and a theoretico-computational framework (DDM) in order to investigate the associations between cognitive processing components and spatio-temporal characteristics of top-down modulations introduced by language during ambiguous visual perception.

Methods

Subjects

Twenty-seven volunteers (13 males) recruited via the Radboud Research Participation System participated in the MEG experiment. The study was approved by the local ethics committee (CMO Arnhem-Nijmegen, Radboud University Medical Center) and conducted in accordance with their guidelines. We made a final selection of subjects based on averaged performance across 2 sessions. We included subjects who managed to withdraw their response to the catch trials in at least 60% of the cases (see details about the procedure below). We excluded 4 subjects due to this issue. The final sample consisted of 23 subjects (Dutch native speakers, 10 males, right-handed, age-range: 22-30, mean age = 25.73, *sd* = 2.45). Another pool of volunteers performed the semantic ratings of priming words (more information below). All participants reported no neurological disease, and had normal or corrected-to-normal vision. All participants signed informed consent forms before the experiment and received monetary compensation for their participation.

Stimuli

For the target stimuli, we used 20 pictures of faces and 20 pictures of houses. We created realistic 3D gender-morphed faces with FaceGen Modeller 3.5 (Singular Inversions). The technical details on the computation method used by the software are discussed in Blanz and Vetter (1999). The face stimuli were presented frontally and cropped using a rectangular shape around the oval of the face. For the house stimuli, we took pictures of the local houses in Nijmegen. All images were cropped to 145 × 155 pixels, placed in a transparent plane of 400 × 400 pixels and further displayed on the screen with the resolution of 1920×1080. We controlled for luminance using SHINE toolbox for MATLAB (Willenbockel et al., 2010).

We preselected prime words using a combination of scores obtained from the word2vec computational model (Mikolov, Chen, Corrado, & Dean, 2013) and naïve subject ratings. First, we extracted the cosine similarity measure between a word-candidate and the target concept 'face' or 'house'. The higher the cosine similarity, the closer the word-candidate is related to the target concept. We preselected 40 words with the highest cosine similarity for the target face concept (face-related words) and 40 with the highest cosine similarity for the target house concept (house-related words). We also preselected 40 words which were further away from the target concept of a face or a house. The words were additionally rated by naïve participants (26 subjects, 11 males, right-handed, age-range: 21-67, mean age = 28.11, *sd* = 9.72). In the rating experiment, the participants had to indicate for each word how related it was to the word "face" or "house" on a 3-point scale (1 = related to house; 2= not related to either face or house; 3 = related to face). For half of the participants, house and face axes were swapped. Based on the rating outcomes, a selection of 60 words was made, which included 20 house-related and 20 face-related words and 20 neutral words (see Supp.Table 1). House-related words were closest to the concept of 'house' in comparison to face-related words and neutral words: house-related > face-related words, (t (19) = 12.24, p < .001); house-related > neutral words, (t (19) = 19.57, p < .001). Face-related words were closest to the concept of 'face' in comparison to house -related words and neutral words: face-related > house-related words, (t (19) = 10.54, p < .001); face-related > neutral words, (t (19) = 20.69, p < .001). Neutral words were furthest away from the target concept of house (target: 'house', face-related > neutral words, t (19) = 6.86, p < .001) and face (target: 'face', house-related > neutral words, t (19) = 7.04, p < .001). Face- and house associated words were matched for word length, frequency per million and concreteness (all *p* > .16). Frequencies for all words were extracted using Subtlex corpus (Keuleers, Brysbaert, & New, 2010). Mean frequency, concreteness, and length are indicated in Table 1.

Table 1 Characteristics of the words used in the priming experiment.

Related word groups	Length	Frequency	Concreteness
house (N = 20)	M = 6.7 (SD = 2.40)	M = 3.87 (SD = 4.62)	M = 4.78 (SD = 0.11)
face (N = 20)	M = 7.25 (SD = 2.02)	M = 5.02 (SD = 6.61)	M = 4.70 (SD = 0.14)
neutral (N = 20)	M = 7.80 (SD = 2.19)	M = 2.37 (SD = 2.70)	M = 4.74 (SD = 0.12)

Procedure

The experiment consisted of 2 phases: the preparatory and the main session spread over 2 separate days. The preparatory session was set to determine the maximally ambiguous target stimuli to be used in the ambiguous trials. Participants underwent a one-up one-down staircase procedure (Cornsweet, 1962). During the staircase procedure we presented simultaneously two pictures (face, house) superimposed on top of each other. We used 20 picture pairs of faces and houses that we produced in a manner similar to the ones used in the main experiment (see Stimuli).

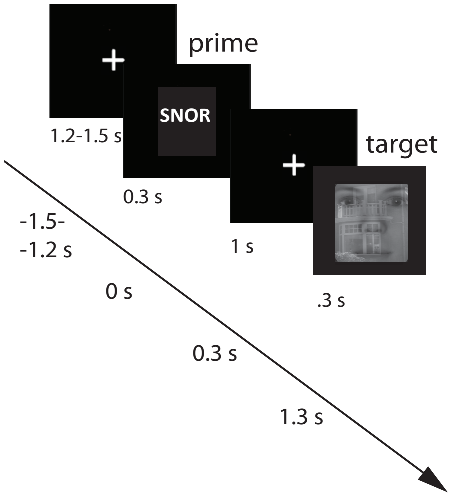


Figure 1: Trial structure timeline with its respective presentation times (next to the picture). For all the analyses we consider the prime onset as 0, pre-target period was the onset of the prime till the onset of the target (0 – 1.3 s), post-target period defined as the onset of the target + .6 s (1.3 – 1.9 s).

For each pair of pictures, we fixed the transparency value of one picture and adjusted the transparency value of another picture upon the reversal (more about reversals in Cornsweet, 1962). There were maximum 8 reversals with decreasing step size on each reversal. The step sizes were: 55, 50, 44, 39, 24, 15 (transparency level, i.e., alpha set in Presentation Neurobs). We discarded the first four trials and averaged the transparency level over the reversals. These pictures-pairs with defined transparency levels for each pair were further used in the main experiment.

In the main experiment, each trial started with a 1.2-1.5 s jittered fixation cross followed by a word (house-related, face-related or neutral) presented for .3 s, which was followed by a fixation cross (1 s) and a target picture (face, house or an ambiguous representation of both) presented for .3 s (Figure 1). Participants were asked to make a binary decision assigning the target into face or house categories. To ensure that participants read the prime word, we introduced “catch” words and instructed participants to skip trials containing pseudowords. These catch trials were excluded from subsequent analyses. We instructed participants to blink after the presentation of a target picture. Participants replied with a bitsi-box button placed in the MEG room by pressing the middle or index finger of the right hand. The buttons were counterbalanced across subjects. The main experiment consisted of 2 sessions recorded on separate days. Each session consisted of 465 trials (300 ambiguous trials, 120 fillers and 45 catches). Ambiguous trials had an ambiguous picture as a target (100 per each word type), filler trials had a non-ambiguous picture as a target (20 trials per each word type). The order of stimuli was randomized across sessions and participants. We used Presentation software (version 17.1, www.neurobs.com) in order to display the stimuli during the experiment.

Behavioural analyses

We excluded 4 subjects who did not manage to correctly withdraw their response in the catch trials. We analyzed reaction times (RT) and percentage of responses for trials with ambiguous targets excluding from the analysis catch trials and filler trials. For trials with ambiguous targets, we performed a repeated measures ANOVA with response (face, house) and prime type (face-related, house-related or neutral) as within-subject factors. We further followed-up this analysis with t-tests to assess the directionality of the effect separately for pictures judged as faces and houses. In addition, we performed a congruency-with-response analysis where we contrasted congruent (ambiguous pictures perceived as faces with face-related primes and ambiguous pictures perceived as houses with house related primes) vs. incongruent (ambiguous pictures perceived as faces with house-related primes and ambiguous pictures perceived as houses with face related primes) vs. neutral (ambiguous pictures perceived as faces or houses with neutral primes) prime-target responses. The analyses were performed using JASP (JASP Team, 2018).

Modelling analyses: Hierarchical Drift Diffusion Model (HDDM)

We analyzed choice reaction-time data with the hierarchical drift-diffusion model, implemented in the Python toolbox HDDM 0.6.0 (Wiecki et al. 2013). In the hierarchical Bayesian version of DDM (HDDM), the model parameters are estimated simultaneously both at the subject and group levels. This in turn allows for enhanced statistical power during the fitting routines since fewer trials are required to recover the parameters. Furthermore, the estimated parameters are less susceptible to outliers (Wiecki, Sofer, & Frank, 2013) making the HDDM an ideal analytic approach for the present study.

We started the HDDM analysis with a theory-free model (all parameters free to vary over all the levels of the experimental condition) since we did not have any a-priori hypotheses on which (if any) HDDM parameters would capture the congruency effect. Next, we contrasted the theory-free model with a number of alternative HDDM models. Each alternative model had different number of parameters that were free to vary across the *congruency* factor. Models with all possible combinations of free parameters across the *congruency* factor constituted the model space (see overall list of models in the Supp.Table 2). Models with different combinations of free parameters were fit to the data via Markov Chain Monte Carlo (MCMC) fitting routines. The comparison of different models was carried out by computing the DIC score for each model and then selecting the model with the lowest DIC value as the best Deviance Information Criterion (DIC) fitting model. DIC is a measure of the goodness of fit of the model to the data that penalizes for the number of free parameters to be estimated from the data. Therefore, a model with a lower DIC score is to be preferred over an alternative model with a higher DIC as the most parsimonious explanation of the data. A difference of 10 between two models indicates substantial evidence in favor of the winning model (Liang & Luo, 2020; Zhang & Rowe, 2014). Coding of the ambiguous faces was done associating the response “house” to the upper boundary and response “face” to the lower boundary. For each model, we evaluated the rate of convergence of the numerical fitting routines and then the ability of the model to capture the observed RT distributions. Models that failed to reach convergence or failed to capture the observed RT distributions were excluded from further analyses. Last, we analyzed the parameter estimates of the best fitting model in a rm-ANOVA with congruency as a within-subject factor. The analyses were carried out using JASP (JASP Team, 2018).

MEG data acquisition

MEG data were recorded using a 275-sensor system (CTF Systems Inc., Port Coquitlam, Canada) at a sampling frequency of 1200 Hz. Participants were seated in a dark magnetically shielded room. Three coils, one in each ear and one on the nasion, were used to determine head position relative to the sensors. Head motion was monitored during the experiment using a real-time head localizer (Stolk, Todorovic, Schoffelen, & Oostenveld, 2013). When head motion exceeded 5 mm, subjects were asked to reposition their head to the original location, making use of the visual representation of initial coil position implemented by the real-time localizer. We recorded the position of the subject’s head in the beginning of session 1, which was used as a reference to align to in session 2. This allowed further combination of the two sessions. No post-hoc correction of head motion was performed. A continuous bipolar electrooculogram (EOG) was recorded for offline rejection of artifacts related to eye movements. We used four electrodes around the eyes – one below and above the left eye for vertical EOG, as well as on the temples for horizontal EOG. The ground electrode was placed on the mastoid.

MEG data preprocessing

Data were analyzed using MATLAB version 7.9.0, R2018b (The Mathworks Inc., Natic, MA) and FieldTrip (version 20200301), an open source Matlab toolbox for the analysis of neuroimaging data (Oostenveld, Fries, Maris, & Schoffelen, 2011). Data were defined ranging from -0.5 s before the prime onset and 0.6 s after the target onset. MEG time courses of these trials were visually inspected, and trials that contained artifacts resulting from SQUID jumps and muscle contractions were rejected. Further we inspected visually for malfunctioning MEG sensors, which were excluded for the analyses in case of observed noise. We demeaned the signal and combined the two sessions per each subject. Further we applied a low-pass filter at 100 Hz. We down-sampled the data to 300 Hz to reduce memory and CPU load. Further we performed independent component analysis (ICA) on the MEG data. We browsed through the first 60 components and identified and excluded from the data the eye-movement related components. Finally, we performed an additional visual check to control for the outliers channel- and trial-wise. All these steps were performed blind for the condition the trials belonged to.

MEG data analysis

For interpretability purposes, we calculated the planar gradients and performed the sensor-level analysis on them (Bastiaansen & Knösche, 2000). We calculated event-related fields (ERFs) at the level of sensors. Given our hypotheses regarding the ERFs, we constrained the analysis to a time window of 1.3 to 1.9 s relative to prime onset, corresponding to .6 s of target onset, and left sensors to investigate the N400 effect (Kutas & Federmeier, 2011; Lau et al., 2008). For the M170, we used the same time window but constrained the analyses to occipital sensors (Liu et al., 2002). For time-frequency representations of power (TFRs), we constrained the analysis to a pre-target period of 0 to 1.3 s and post-target period of 1.3 – 1.9 s for ERFs and of 1.3 – 1.8 for TFRs (see motivation below). The frequency range was constrained to 2-15 Hz given the exploratory part of this analysis. We used a pre-prime period of -0.5 to -.05 for the baseline for both ERFs and TFRs analyses.

For ERFs, we performed ANOVA (F-stats, permutation test) to investigate main effect of primes separately for face responses (FF, HF, NF) and for house responses (HH, FH, NH) on the post-target time-interval (1.3 to 1.9 s) for occipital, left and all MEG channels. We followed up each analysis of main effects with a t-test in case of significance. Further, we performed a congruency-with-response analysis (congruent: ambiguous pictures perceived as faces with face-related primes and ambiguous pictures perceived as houses with house-related primes; incongruent: ambiguous pictures perceived as faces with house-related primes and ambiguous pictures perceived as houses with face related primes, neutral: ambiguous pictures perceived as houses or faces with neutral primes), which was followed-up with t-tests in case of significance. We performed the congruency analysis separately for occipital, left and all MEG channels.

For TFRs, we performed the analyses similar to ERFs (main effect of primes and congruency-with-response), adding to post-target (1.3 – 1.8 s) testing an exploratory testing of pre-target period (0 – 1.3 s). We note that we tested post-target 1.3 – 1.8 s instead of 1.3 – 1.9 s due to the available frequency resolution. We performed statistical testing based on a non-parametrical cluster-based permutation technique (Maris & Oostenveld, 2007). We used 1000 permutations for the analysis, which is a recommended number of permutations (Pernet, Latinus, Nichols, & Rousselet, 2015).

Correlation analysis

We planned to perform a Spearman correlation test between the HDDM parameters and neural markers of priming (if any) with the purpose of testing whether neural markers of language bias can be associated with one of the parameters derived from the formal model of decision making (i.e. decision threshold or drift rate). The analyses were carried out using JASP (JASP Team, 2018).

Results

Behavioural

Results for the RT and percentage of F/H responses are summarized in Figure 2.

For the responses (Figure 2, A), we did not find that any of the words biased the house-face response more than the others (main effect of prime: *F* (2, 44) = 2.91, *p* = 0.065; main effect of response: *F* (1, 22) = 4.06, *p* = 0.05; interaction between prime and response: *F* (2, 44) = 2.26, *p* = 0.11).

For RTs (Figure 2, B), we found that participants were generally faster for judging images as faces vs. houses (main effect of response: *F* (1, 22) = 5.73, *p* = 0.02). However we did not find that any of the words biased the response towards face or house (main effect of prime: *F* (2, 44) = 1.23, *p* = 0.30: interaction between prime and response: *F* (2, 44) = 2.78, *p* = 0.073).

While we found a main effect of congruency in RTs (Figure 2, D): *F* (2, 44) = 3.27, *p* = 0.047, we did not find any difference in percentage of responses (Figure 2, C): *F* (2, 44) = 1.26, *p* = 0.291. We found that participants were faster for congruent in comparison to incongruent (t (22) = -2.0, p = .028), and for congruent in comparison to neutral (t (22) = -2.52, p = .01) stimuli. We did not find a difference between incongruent and neutral word-target pairs (t (22) = -.026, p = .980).

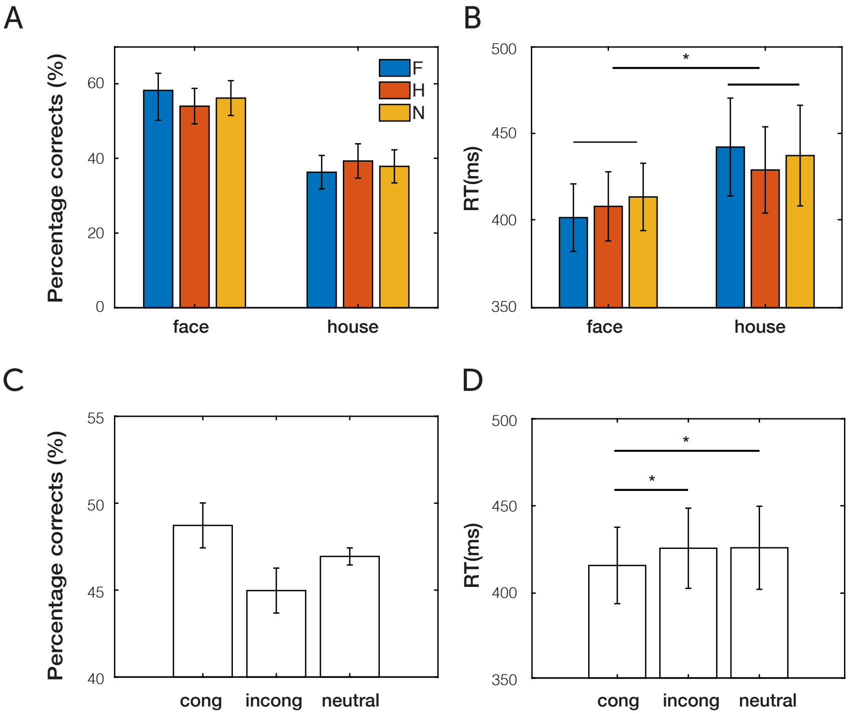


Figure 2: Percentage of responses and response times for ambiguous trials. Bars are standard errors (SE). **(A-B)** On the x axis, face/house are responses. F- face related words, H – house related words, N – neutral words. **(C, D)** 'cong' = prime F, response F and prime H, response H; 'incong' = prime F, response H and prime H, response F; 'neutral': prime N, response F and prime N, response H. Error bars are standard error of the mean. * $p < .05$.

Modelling analyses

Further we conducted the drift diffusion analysis using RTs and choice responses for congruent vs. incongruent conditions.

Model Convergence and Model Fit

Estimation of the model parameters was done via Monte-Carlo Markov Chain (MCMC) simulations. For all of the analyses reported, the MCMC (Gelman & Rubin, 1992) fitting routines were run for 20.000 iterations with a burn-in period of 10.000 interaction and a thinning of 5. Model convergence was assessed by examination of the density estimates and of the R-hat statistic, which are measures of convergence among multiple MCMC chains (three for the present study). Posterior density estimates, which are stable over multiple samples, indicate that the fitting routines have converged to a fixed estimate. For the R-hat statistic instead, a value below 1.1 indicates that chains with different starting values have converged to the same posterior estimate. Successful convergence was furthermore confirmed by a MCMC error smaller than 0.01 for all of the parameters. As an extra check of the quality of the model predictions, we compared observed and predicted RT distributions produced by the model (see Supp. Figure 1 for an example of model fit for congruent and incongruent conditions in the drift rate). After assessing convergence, we carried out quantitative comparisons of alternative models by comparing the associated DIC scores for each model. The model that best described the data (i.e., the model with the lowest DIC score, see Supp. Table 2 for details) was the model with the following parameters estimated per subject: drift rate (*v*), non-decision time (*Ter*), and decision threshold (*A*) free to vary over *congruency* (congruent and incongruent conditions). Conventionally, a DIC difference of more than 10 indicates that the evidence in favor of the model-winner is substantial (Burnham, Anderson, & Burnham, 2002). The second best model (model 1, DIC -16462.9) was the model with decision threshold (A) parameter free over the congruency condition. However since the difference between model 6 and model 1 did not exceed 10 scores, we consider model 6 as the most parsimonious account of the data and therefore further analyses focus on this model.

Model parameter analysis

We performed a rm ANOVA analysis with congruency as within subject factor. Results are summarized in Figure 3.

We found a main effect of congruency in decision threshold (A), F(1, 22) = 14.35, p = .001; in drift rate (V), F(1, 22) = 15.60, p < .001; and in non-decision component (Ter), F(1, 22) = 7.84, p = .010. While decision threshold and drift rate were increased for the congruent condition in comparison to incongruent, the non-decision component was increased for incongruent in comparison to the congruent condition: A (con > inc): t (22) = 3.78, p < .001, V (con > inc): t (22) = 3.95, p < .001, Ter (inc > con): t (22) = 2.80, p = .005.

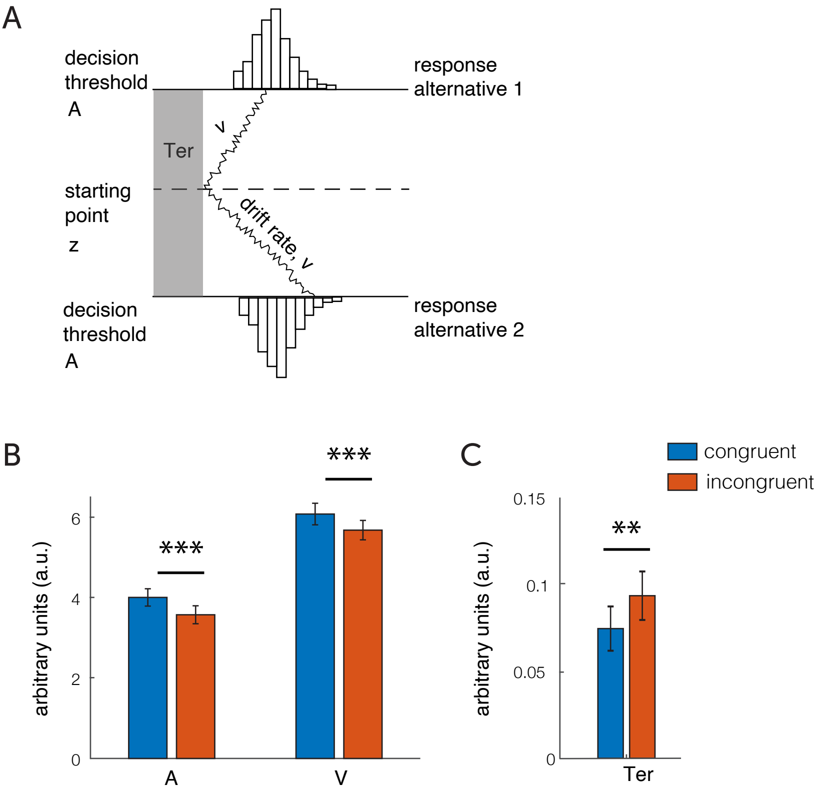


Figure 3: (A) Schematic representation of the drift diffusion model. Once the stimulus is encoded, the process of evidence accumulation begins. The evidence is accumulated towards one of the two decision boundaries. After the boundary is reached, the associated response is executed. The Drift Diffusion model with the four parameters: drift rate (*v*), decision threshold (*A*), starting point (*z*), non-decision component (*Ter*). HDDM parameters: decision threshold, drift rate **(B)** and non-decisional component **(C)** for congruent vs. incongruent conditions. Error bars, standard error of the mean. ** $p < 0.01$, *** $p < 0.001$

Neural results

ERF

We found that primes (face-, house-related or neutral) did not bias neither face responses (F-test for FF, HF, NF) in occipital channels: p = .35, left channels: p = .76; all channels: p = .25, nor house responses (F-test for HH, FH, NH) in occipital channels: p = .08, left channels: p = .11; all channels: p = .20.

We further investigated the response introduced by words (i.e. congruency analysis, see Figure 4). We found significant differences in neural signal for congruent vs. incongruent in *all channels* (p = .03), but not in *occipital channels* (p > .22) nor in *left channels* (p = 1). To sum up, we found statistical differences between congruent and incongruent conditions, the differences were most prominent in the time range of 1.59-1.67s across right occipito-parietal channels.

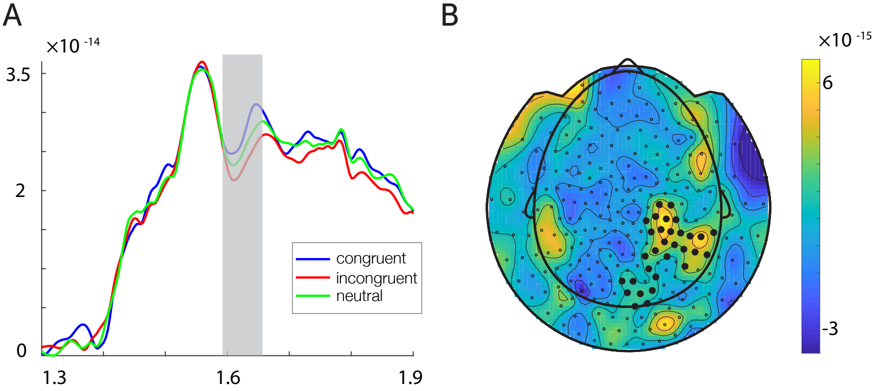


Figure 4: (A) Time-course for congruent and incongruent conditions starting at target picture onset (1.3 s) until the end of the trial (1.9 s). The grey shaded area represents most prominent differences associated with the cluster. **(B)** Topographical representation of the differences between the congruent vs. incongruent conditions, within the time-window of 1.59 to 1.67 s.

TFR

We did not find a main effect of prime either for face responses (0 to 1.3 s: p = .68; 1.3 to 1.8 s: p = .10), or for house responses (0 to 1.3 s: p = .12; 1.3 to 1.8 s: p = .38).

For the congruency analysis, within the oscillatory range of 2-15 Hz no differences were found either in the post-target window (p > .22) or in the pre-target window (p = .48).

Correlational analysis

Results of the Spearman correlation indicated that there was no significant association (positive or negative) between congruent vs. incongruent ERF amplitude and DDM measures. The results of the correlational analysis are summarized in the Supp. Table 3.

Discussion

In this study we investigated the neural effects of top-down modulations introduced by the lexical semantics of words on ambiguous visual perception. Particularly, we tested whether these modulations occur at the level of early visual processing or later stages of processing related to semantic memory or decision making. Participants in our experiment had to decide whether they saw a face or a house while words related to face, house or unrelated to either of them were briefly flashed before the ambiguous target image. We found that primes congruent with the response given (house-related words resulted in house responses and face-related words resulted in face responses) led to faster RTs in comparison to primes incongruent with the response given (house-related words resulted in face responses and face-related words resulted in house responses). In the neural data we found larger ERF amplitude for congruent in comparison to incongruent trials around 300 ms post-target onset.

The primary goal of this study was to clarify which process mediates the effect of words on ambiguous perception – early visual or late semantic. We did not find either neural markers related to lexico-semantic processing (N400m), or to early visual processing (M170), which suggests that based on this experimental evidence, and given the statistical power limits of our study, these processes are not at play while participants are deciding about the category of an ambiguous picture. Further, we tested whether words can elicit pre-stimulus activity in TFRs. This was motivated by previous research that has found that posterior alpha power reflects the pre-activated knowledge about an object's identity when participants have a meaningful cue that they use for the interpretation of upcoming ambiguous faces (Samaha et al., 2018). In our experiment we did not find pre-stimulus alpha oscillations, which suggests that in the word-picture priming paradigm with a semantic categorization task, words are unlikely to induce pre-stimulus top-down influence. This might be due to aspects of our experimental design such as a relatively long word-target interval (1s) or the task itself (classification, rather than matching, for example). Future research might make use of multivariate techniques that have been suggested to be useful in detecting sharpening of stimulus representations undetectable with univariate methods (Kok, Jehee, & de Lange, 2012). Contrary to our predictions, task-related words (face-related and house-related words) that were congruent with the response resulted in a larger ERF amplitude that peaked around 300 ms in comparison to task-related words that were incongruent with the response. There might be two potential interpretations of this finding. First, this effect can be interpreted as the N300 component, which has been shown to index the speed of matching the visual input to stored semantic knowledge (Schendan &

Kutas, 2007; Sitnikova, Holcomb, Kiyonaga, & Kuperberg, 2008). It has been shown that when the word and the picture are congruent then the N300 amplitude is attenuated in comparison to the word and the picture that are incongruent (Yum, Holcomb, & Grainger, 2011). In line with these previous findings, the results of the current experiment suggest that congruent words promote the ability to match a semantic representation to the visual input, which is reflected by a large N300 Congruency response (higher for congruent vs. incongruent items). While the results of the current study suggest that the topographical representation of the differences between the congruent vs. incongruent conditions in rather in occipi-to-parietal areas, previous studies report a frontal topography of the N300 effect. The frontal topography was derived on the basis of scalp electroencephalography (EEG) data, which might partially explain the difference in topographies. However this discrepancy triggers us to further consider alternative interpretations to the N300 interpretation.

Another interpretation of the ERF amplitude differences that peaked around 300 ms points to the P300 component. The P300 is an event-related component that appears at 300-600 ms following any task-related stimulus, regardless of its modality (Nolan et al., 2012; Peng, Hu, Zhang, & Hu, 2012; Zaslansky, Sprecher, Tenke, Hemli, & Yarnitsky, 1996). Due to its relevance in many cognitive operations, the function of this component is still unclear (Linden, 2005). Mostly it has been associated with the processes of attention and memory updating (Comerchero & Polich, 1999; Polich, 2007), but also with decision confidence (Hillyard, Squires, Bauer, & Lindsay, 1971) and surprise (Mars et al., 2008). Research in the field of semantic integration has shown that the P300 can be correlated with aspects of semantic access (Dorjee, Devenney, & Thierry, 2010; Hill, Ott, & Weisbrod, 2005; Hill, Strube, Roesch-Ely, & Weisbrod, 2002; Rossell, Price, & Nobre, 2003). Particularly, P300 with a temporo-parietal topography has been associated with memory retrieval (Polich, 2007) and the decrease of its amplitude has been associated with increased memory load (Mecklinger, Kramer, & Strayer, 1992). Thus, it could be that decreased P300 reflects enhanced memory load for incongruent in comparison to congruent word-picture pairs. To conclude, it is a matter of future investigations, with potential use of source localization techniques, to clarify whether the effect found in this study constitutes a P300 or N300 effect. While being an interesting question, answering it immediately is not crucial for the purpose of this study since we focused more broadly on whether cognitive markers of language-vision integration can be associated with the HDDM parameters.

We further aimed at elaborating on the cognitive mechanisms of the P300/N300 component. Particularly, we considered language-vision integration within the theoretico-computational framework of the sequential sampling models (e.g. DDM). The goal of this theoretical consideration was to draw a parallel between evidence accumulation and the P300/N300 component observed. Integration is an important property of decision formation and in sequential sampling models it is characterized as the noisy accumulation of evidence (information) over time. As discussed by Summerfield (Summerfield & Tickle, 2015, p.1635) "sensory inputs are integrated by summation until a threshold is met, at which point a response is initiated". Previous animal studies convincingly showed that this process is mediated by increasing neuronal firing rates in cortical neurons, which scales with information quality. Information quality is an abstraction that can reflect, for example, properties of the stimulus (more degraded stimuli have poorer information quality), or speed of the decision (words with unrelated primes in comparison to word with related primes are of poorer information quality). In human experimental work, a recent study has shown that P300 amplitude scaled with the degree of information quality (Twomey, Murphy, Kelly, & O'Connell, 2015) using a combination of mathematic modeling and EEG. Another EEG study investigating decision processes during degraded visual categorization showed that the late effect (~300 ms post target) correlated with the rate of evidence accumulation, highlighting the role of decision rather than perception process during degraded categorization. In this study we correlated P300/N300 amplitude with the computational parameters of the DDM that characterized the process of information integration in this task (i.e. congruency effect). We did not find associations between the HDDM parameters and P300/N300. There are several explanations for that. First, it can be that inter-trial variability in evidence accumulation (drift rate) plays a critical role in the process of integration of language and vision information. For example, neural activity before the language prime can also affect the decision and response about the ambiguous target. To account for this, trial-by-trial variability needs to be added to the model. However in the present study this was not possible due to the larger number of trials required for such an analysis. Second, it can be that the ERF effect in this task, rather than reflecting aspects related to decision making, reflects other mechanisms such as confidence or surprise. Further investigations are needed in order to clarify whether the language advantage in ambiguous visual perception can be connected to the notion of decision making such as enhanced information quality. Finally, it is important to note the analytical limitations. Here we performed only ERF and TFR analyses, disregarding other analytical options. In other words, there might be other neural markers that could reflect the relationship between language and decision making in visual decisions. To discover these markers, new analytical pipelines and approaches are needed. For example, it can be that the interaction between oscillations in different frequency bands (or cross-frequency coupling) can be more informative in understanding language-perception interactions. For example, it has been shown that phase-amplitude coupling (a type of cross-frequency coupling) is potentially involved in attentional selection, memory retention and sensory integration (Lisman, 2005; Lisman & Idiart, 1995; Schroeder & Lakatos, 2009). It would be an intriguing next step to apply the methods of cross-frequency coupling to investigate language-perception interactions.

To sum up, we found that effect of words on ambiguous visual perception in a forced choice binary task involves post-perceptual processes associated with attention or decision making, rather than the engagement of processes related to early visual perception or lexical-semantic memory.

5

Lexical-semantic and executive deficits revealed by computational modelling: a drift diffusion model perspective

A modified version of this chapter has appeared as: Todorova L., Neville D.A., Piai V. (2020). Lexical-semantic and executive deficits revealed by computational modelling: a drift diffusion model perspective. *Neuropsychologia*, 146.

Abstract

Flexible language use requires coordinated functioning of two systems: conceptual representations and control. The interaction between the two systems can be observed when people are asked to match a word to a picture. Participants are slower and less accurate for related word-picture pairs (word: banana, picture: apple) relative to unrelated pairs (word: banjo, picture: apple). The mechanism underlying interference however is still unclear. We analyzed word-picture matching (WPM) performance of patients with stroke-induced lesions to the left-temporal (N = 5) or left-frontal cortex (N = 5) and matched controls (N = 12) using the drift diffusion model (DDM). In DDM, the process of making a decision is described as the stochastic accumulation of evidence towards a response. The parameters of the DDM model that characterize this process are decision threshold, drift rate, starting point and non-decision time, each of which bears cognitive interpretability. We compared the estimated model parameters from controls and patients to investigate the mechanisms of WPM interference. WPM performance in controls was explained by the amount of information needed to make a decision (decision threshold): a higher threshold was associated with related word-picture pairs relative to unrelated ones. No difference was found in the quality of the evidence (drift rate). This suggests an executive rather than semantic mechanism underlying WPM interference. Both patients with temporal and frontal lesions exhibited both increased drift rate and decision threshold for unrelated pairs relative to related ones. Left-frontal and temporal damage affected the computations required by WPM similarly, resulting in systematic deficits across lexical-semantic memory and executive functions. These results support a diverse but interactive role of lexical-semantic memory and semantic control mechanisms.

Introduction

Flexible language use requires coordinated functioning of two systems: conceptual representations and control (Lambon Ralph, Jefferies, Patterson, & Rogers, 2017). The system of conceptual representations entails a distributed network of cortical regions that code information multi-modally (Barsalou, 2008; Warrington & Shallice, 1984) and allow to generalize to new knowledge thanks to already existing representations (Patterson, Nestor, & Rogers, 2007; Rogers et al., 2004). The control system manipulates the semantic knowledge depending on the task at hand (Hoffman, Lambon Ralph, & Rogers, 2013; Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997). Understanding the contributions of control mechanisms to lexical-semantic processes is an important theoretical question, with potential applications for clinical research.

Control mechanisms (henceforth referred to as semantic selection) operate over several automatically activated semantic representations to ensure the functioning of semantic cognition (Lambon Ralph et al., 2017). Multiple co-activated semantic representations can result in competition depending on the semantic selection demands in the context of a task (Chiou, Humphreys, Jung, & Lambon Ralph, 2018; Jefferies, Rogers, Hopper, & Lambon Ralph, 2010). Competition between (lexical-) semantic representations can manifest itself via semantic interference in tasks such as word-picture matching (Campanella & Shallice, 2011; Crutch & Warrington, 2005) or picture-word interference (Lupker, 1979; Piai & Knight, 2018). Semantic interference is reflected in poorer performance when comparing semantically related word-picture pairs to unrelated ones. For example, participants are slower when they have to match a picture preceded by a related word (word: "banana", picture: apple) as opposed to an unrelated word (word: "banjo", picture: apple). While various psycholinguistic studies propose theoretico-computational accounts of interference effects in language production using picture naming paradigms (Howard, Nickels, Coltheart, & Cole-Virtue, 2006; Oppenheim, Dell, & Schwartz, 2010; Roelofs, 2018, 1992), mechanistic explanations of this effect in language comprehension using word-picture matching types of tasks are scarce. Most of the computational models in the language production literature that could be extended to explain semantic interference in word-picture matching are neural network models that operate at the level of the excitatory-inhibitory connections between conceptual and lexical nodes (Howard et al., 2006; Oppenheim et al., 2010; but see Roelofs, 2003). It is generally accepted that in addition to lexico-semantic processes, executive mechanisms govern the extraction of relevant semantic information from stimuli (Badre, Poldrack, Paré-Blagoev, Insler, & Wagner, 2005; Jefferies & Lambon Ralph, 2006; Lambon Ralph et al., 2017; Nagel, Schumacher, Goebel, &

D'Esposito, 2008). The exact mechanisms – lexico-semantic and/or control – governing retrieval and selection in competitive semantic contexts are still not clear.

In this study, we investigated the task demands elicited by competing semantic representations using a word-picture matching paradigm, in which participants verify whether a word and a subsequently presented picture match. Behavioral performance was then modelled using the drift diffusion model (DDM, (Ratcliff, 1978), an analytical approach that combines reaction time distributions for correct and incorrect task responses in order to estimate latent variables associated with task performance. According to the DDM framework, any binary decision-making task is described as a stochastic accumulation of evidence over time towards one of the two decision boundaries (Figure 1). Once the decision is reached, then the response associated with that boundary is produced (Voss, Nagler, & Lerche, 2013). In the DDM, a total of four parameters describe the components underlying decision making: the rate with which evidence accumulates over time (*drift rate*), the amount of evidence that is necessary to reach a response (*decision threshold*), the amount of information accumulated before the decision process has started (*starting point*) and the time required by *non-decision processes*, such as motor preparation or stimulus encoding. Moreover, we applied the DDM approach not only to decision-making data from neurotypical participants, but also to data from individuals with frontal or temporal lobe lesions. The lesion approach can help further clarify the relationship between the latent variables and the presupposed cognitive levels. In addition, we used a hierarchical version of DDM (HDDM) where model parameters were estimated in a Bayesian framework, allowing for accurate model fits even with relatively few trials. This approach provides a great advantage for the analysis of data from clinical populations, where collecting thousands of experimental trials for an accurate model fit (Brunton, Botvinick, & Brody, 2013) is unfeasible.

Previous psycholinguistic experiments that applied drift diffusion formalism showed that the drift rate reflects processes operating at the level of lexical semantics. For example, a larger drift rate was found for words in comparison to random letter strings but not in comparison to word-like non-words (Ratcliff, Gomez, & McKoon, 2004). Moreover, words of higher frequency had a higher drift rate compared to words of lower frequency (Ratcliff, Gomez, et al., 2004). Semantic priming experiments with lexical decision brought additional evidence that the drift rate serves as a proxy for processes related to lexico-semantic representations (Meyer & Schvaneveldt, 1971; Neely, 1991; Wentura, 2000). Particularly, during lexical decision, related word-target pairs (word: "lion", target: tiger) resulted in a higher drift rate as opposed to unrelated targets (word: "king", target: bee). Interestingly, the results remained

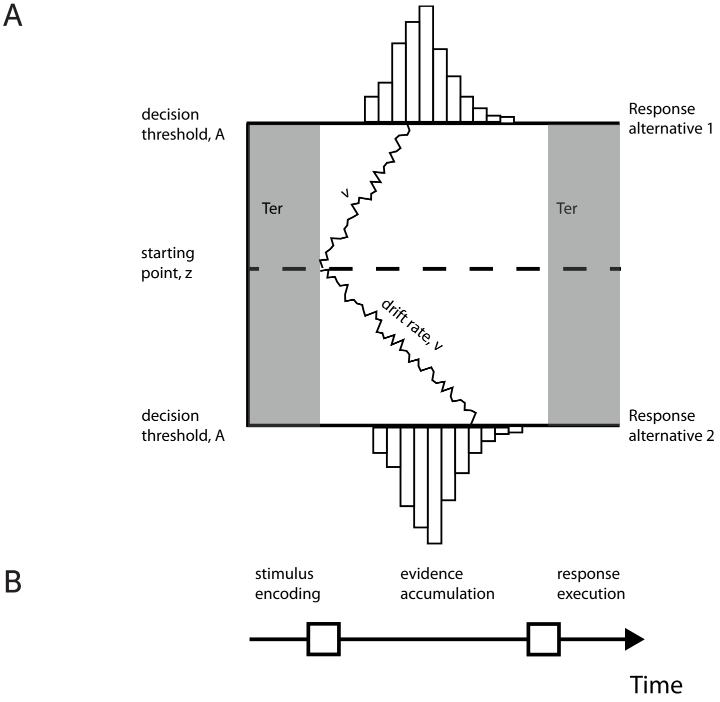


Figure 1: Schematic representation of the drift diffusion model. Once the stimulus is encoded, the process of evidence accumulation begins. The evidence is accumulated towards one of the decision boundaries. After the boundary is reached, the response is executed. We adopted the following notation for the model parameters description: non-decisional component (Ter), drift rate (v), decision threshold (A), starting point (z).

similar even when changing the task from lexical to semantic decision, i.e., when participants had to decide if the target was a living entity (Voss, Rothermund, et al., 2013). This finding was interpreted in terms of semantic facilitation of lexical access. The fact that lexicality, word-frequency, and semantic priming effects are reflected in the drift rate is an indication that the drift rate reflects lexical-semantic processes.

In the field of decision making, it has been consistently shown that modulation of the decision threshold is directly related to how cautiously people behave (Bogacz, Hu, Holmes, & Cohen, 2010; Forstmann, Brown, Dutilh, Neumann, & Wagenmakers, 2010). For example, if participants prioritize accuracy over speed in their responses then the decision threshold is raised, i.e., it takes more time to accumulate information towards one decision threshold (responses are slow but accurate). The decision threshold is modulated by conflict tasks, being higher for the most conflicting conditions in comparison to the less conflicting ones (Cavanagh et al., 2011). Moreover, in traditional executive tasks, such as the Stroop task, the decision threshold is associated with mechanisms of executive control (Kinoshita et al., 2017).

While the number of studies that applied DDM modelling to experiments investigating aspects of lexico-semantic retrieval is modest, they bring indicative results about the relevance of the decision threshold. For example, Anders and colleagues showed that increasing semantic interference in a blocked-cyclic picture naming task resulted in decrease of lexical accumulation activity (drift rate) and increase in the amount of activation necessary for lexical target selection (threshold) for healthy subjects (Anders, Ries, Van Maanen, & Alario, 2017). Crucially, patients with prefrontal lesions failed to adjust the decision threshold appropriately in comparison to controls (i.e., no up-adjustment for the most interfering condition). Anders et al. highlight that patients with prefrontal lesions have impaired selection processes in word production (Schnur et al., 2009), that can potentially be more domain-general (Ries, Greenhouse, Dronkers, Haaland, & Knight, 2014; Ries, Karzmark, Navarrete, Knight, & Dronkers, 2015). We note that the discussion about the nature of the control mechanisms in the context of language tasks is still ongoing, and it is outside the scope of the present study to contribute to that debate (Hoffman, 2018; Noonan, Jefferies, Visser, & Lambon Ralph, 2013; Thompson-Schill et al., 1997).

Neuropsychological evidence for semantic cognition posits that processes of semantic knowledge and control have different neural substrates. While anterior and ventro-lateral parts of left temporal lobe are crucial for semantic knowledge (Hickok & Poeppel, 2004; Patterson et al., 2007), left prefrontal regions are associated with semantic retrieval (Lambon Ralph et al., 2017; Noppeney, Phillips, & Price, 2004). Indeed, previous studies investigating the role of left prefrontal cortex (PFC) in word selection showed that left PFC helps overcome semantic interference by boosting mechanisms of control (Ries et al., 2014), especially when selection demands are difficult (Ries et al., 2015). However, PFC involvement may be dependent on the task and individual variability (Piai & Knight, 2018; Piai, Ries, & Swick, 2016; Python et al., 2018). Recently, it has been proposed that prefrontal regions, in particular the inferior frontal gyrus (IFG), operate in conjunction with the posterior middle temporal gyrus (MTG) contributing to the mechanisms of semantic control (Davey et al., 2016; Carin Whitney, Kirk, O'Sullivan, Lambon Ralph, & Jefferies, 2012). Understanding the mechanisms of impairment in semantic retrieval and selection is still an ongoing task for the language community.

We investigated the processing stages of semantic interference in healthy participants and patients with left temporal or frontal lesions. We used a word-picture matching task (i.e., participants indicated whether the word and the picture matched), where the picture was preceded by related words (word: "banana", picture: apple), unrelated words (word: "banjo", picture: apple) or words that directly denoted the picture (word: "apple", picture: apple). We had the following hypotheses regarding the experimental results. First, the interference effect can be reflected in the drift rate, which would support the idea that the interference resides at the lexico-semantic level. We would expect the drift rate to decrease for the related pairs as opposed to the unrelated pairs, following the finding that semantic interference in blocked-cyclic naming is reflected in decreased drift rate (Anders et al., 2017). We note that this prediction is contrary to the semantic priming findings (Voss, Rothermund, et al., 2013). Alternatively, or in addition to the effects at the level of evidence accumulation (drift rate), the interference effect can modulate the decision threshold, reflecting the amount of information needed to reach the decision. In this case, we would expect higher threshold for the condition that evokes more interference i.e., related relative to unrelated word-target pairs. We note that this prediction is both in line with Anders et al. (2017), and follows the logic of conflict tasks (Cavanagh et al., 2011; Kinoshita et al., 2017). This would support the idea that executive control is involved in the presence of semantic competition, which is strengthened by the semantic relationship between the picture and the word. After having established the mechanisms of semantic interference in healthy participants, we compared the interference effect between healthy subjects and the individuals with left frontal or temporal stroke-lesions. Due to the fact that patients with temporal lesions had a combination of lesions in both anterior and posterior sites of MTG, we expected a joint effect of semantic competition on the drift rate and decision threshold reflecting deficits at potentially both lexico-semantic and executive levels. For the patients with frontal lesions, we would expect deficits in the decision threshold as it reflects strategic cognitive control allocation (Cavanagh et al., 2011; Domenech & Dreher, 2010). Damage to the frontal cortex might also lead to the deterioration of the activation levels in the lexico-semantic domain (which would be reflected in decreased drift rate) due to the connected frontal and temporo-parietal networks that underlie semantic control (Lambon Ralph et al., 2017).

To sum up, we investigated the contribution of control mechanisms during semantic retrieval and selection in healthy adults and adults with stroke-induced lesions using a word-picture matching task and computational modelling (HDDM). With that, we add to an existing body of literature that combines computational modelling with a lesion mapping approach, which provides a promising avenue towards formalizing brain-behavior relations (Chen, Middleton, & Mirman, 2018; Dell, Martin, & Schwartz, 2007; Hoffman, McClelland, & Lambon Ralph, 2018).

Methods

Participants

Twelve patients with stroke-induced lesions to the left lateral-temporal or lateral-frontal cortex participated (five females; median age = 66, mean = 65, sd = 9, range = 50-74; mean years of education = 17). All patients were English native speakers. One additional patient with Wernicke's aphasia was included, but failed to understand the instructions and testing was therefore discontinued. We differentiate patients with lesions in temporal and frontal lesions based on the topology of the predominant lesions. While patients with temporal lesions had lesions predominantly in temporal areas (MTG, STG) but not in fronto-striatal regions, patients with frontal lesions had lesions predominantly in frontal or striatal areas. We excluded two patients that had lesions in both temporal and fronto-striatal areas. Patients were tested at least 12 months post stroke and were pre-morbidly right-handed. Additionally, 12 right-handed controls participated, matched to the patients for gender (five females), age (within ±4 years of age, median age = 66, mean age = 64, sd = 8, range = 50-74, $t < 1$, $p = .873$), and years of education (±4 years of education, mean years of education = 17, $t < 1$, $p > 0.949$). All participants were native speakers of American English and none had a history of psychiatric disturbances, substance abuse, medical complications, multiple neurological events, or dementia. Information on the patients' lesions and language ability are shown in Tables 1 and 2. The study protocol was approved by the University of California, Berkeley Committee for Protection of Human Subjects, following the declaration of Helsinki. All participants gave written informed consent after the nature of the study was explained and received monetary compensation for their participation.

Lesion analysis

Lesions were drawn on patients' structural magnetic resonance images (MRIs) by a trained technician and confirmed by a neurologist. Lesions masks were then normalized to the MNI template. Percent damage to different areas was determined based on the Automated Anatomical Labeling template in MRIcroN (Rorden & Brett, 2000). The maps of lesion overlap are presented in Figure 2. In the patients with temporal lesions, the lesions were mostly present in the left middle temporal gyrus (100% overlap). In the patients with frontal lesions, the lesion overlap was centered on the left inferior frontal gyrus and left middle frontal gyrus (100% overlap).

Table 1 Individual percent damage to the striatum (joint measures for caudate, putamen and globus pallidus), left inferior frontal gyrus (IFG: opercularis, triangularis, orbitalis), middle frontal gyrus (MFG, including middle and middle opercularis), middle temporal gyrus (MTG, including middle and inferior MTG), superior temporal gyrus (STG) and parietal lobe (POST: Postcentral gyrus, SMG: Supramarginal gyrus, AG: angular gyrus, P2: inferior parietal gyrus, P1: Superior parietal gyrus).

Patient	Striatum	IFG	MFG	MTG	STG	POST	SMG	AG	P2	P1
Left temporal lobe lesions										
P 1	0	0	0	23.6	34.0	.2	12.5	2	0	0
P 2	0.1	0.2	0	31.5	67.4	12.6	85.1	.4	35.9	0
P 3	0	0.1	0	33.4	88.1	4.7	88.3	57.9	25.8	0
P 4	0	0	0	7.9	0.8	0	0	0	0	0
P 5	1.8	0	0	48.1	7.8	0	0	0	0	0
Left frontal lobe lesions										
P 6	81	93.4	53.3	0.1	1.3	4.4	0	0	0	0
P 7	30.53	78.03	37.21	0	10.1	6.1	0	0	0	0
P 8	23.8	57.9	27.9	0	49.8	58.2	71.6	.5	25.7	.2
P 9	0	6.1	7	0	0	2.4	0	0	0	0
P 10	33.3	57.2	4.75	0	12.9	0	0	0	0	0

Table 2 Language testing data from the Western Aphasia Battery (WAB).

	AQ	Naming	Comprehension
P 1	NA	NA	NA
P 2	63.9	7.9	8.55
P 3	92.9	9.5	9.55
P 4	99.6	10	10
P 5	94	8.6	10
P 6	91.6	9.2	10
P 7	92.05	9.3	8.825
P 8	87.2	8.9	8.9
P 9	NA	NA	NA
P 10	99.6	9.8	10

Naming = WAB Naming and Word Finding score (maximum = 10). Comprehension = WAB Auditory Verbal Comprehension score (maximum = 10). Aphasia Quotient (AQ, maximum = 100).

Lesion distributions for each patient group are shown in Figure 2. Five patients had lesions predominantly in the left frontal lobe and five in the left temporal lobe. Language scores (Western Aphasia Battery, WAB, Kertesz, 1982) were available for ten patients.

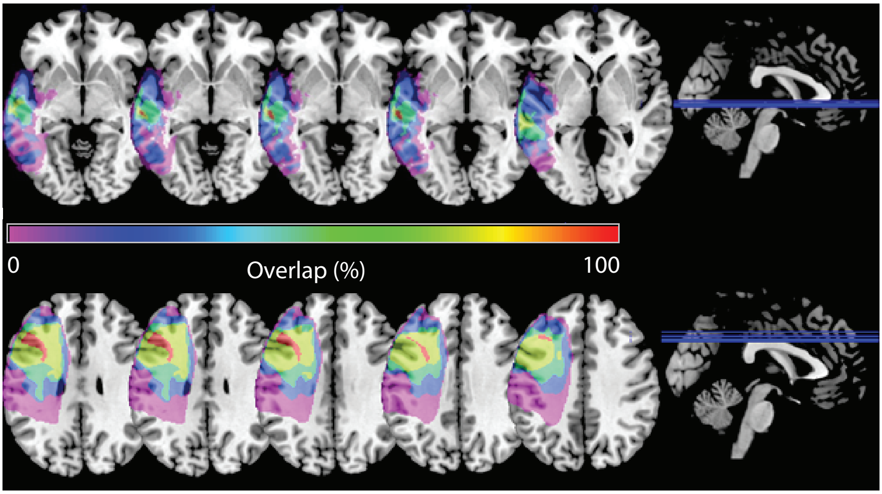


Figure 2: Upper. Lesion overlap of individuals with temporal lobe lesions (100% overlap in left middle temporal gyrus). Lower. Lesion overlap of individuals with frontal lobe lesions (100% overlap in left inferior and middle frontal gyrus).

Materials

Seventy pictures were selected from the BOSS database (Brodeur, Dionne-Dostie, Montreuil, & Lepage, 2010) together with their basic-level names. For each picture, three conditions were created. In the congruent condition, the prime word was the picture's basic-level name. For the semantically related condition, prime words were selected using existing norms (Nelson, McEvoy, & Schreiber, 2004). The forward strength between the prime word and the picture name ranged between 0.108 and 0.879 (mean = 0.451, median = 0.394, sd = 0.2). For the unrelated condition, prime word and target picture names were recombined to form semantically and phonologically unrelated pairs. The association norm in the unrelated condition was 0 for all items. Latent semantic analysis (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990) was also used to confirm the strength in semantic relationship between the prime word and the picture name (related condition: mean = 0.463, median = 0.42, sd = 0.212; unrelated condition: mean = 0.071, median = 0.05, sd = 0.071, $t(69) = 14.05$, $p < .001$). The prime words were spoken by a female native speaker of American English, recorded in a sound-proof booth and subsequently normalized to 77 dB sound-pressure level.

Procedure

Stimulus presentation and response recording were controlled by Presentation® software (Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com). Participants were tested individually in a sound-attenuated, dimly-lit booth. The words were presented via loudspeakers. Participants responded with a left-hand button press, using the index finger for 'yes' responses and the middle finger for 'no' responses.

A trial began with a fixation cross displayed for 1 s. The fixation cross stayed on the screen during word playback and then during a silent period of 1 s between auditory word offset and picture onset. Then the picture was presented for 2 s and participants responded whether the word and the picture were a match (word: apple, picture: apple) or a mismatch (word: banana, picture: apple) during this period. Three asterisks, indicating termination of the trial (***) then appeared for a variable interval between 1.2 and 1.9 s. An example of an experimental trial is given in Figure 3. There were 280 experimental trials (70 for each of the related and unrelated conditions, and 140 for congruent condition).

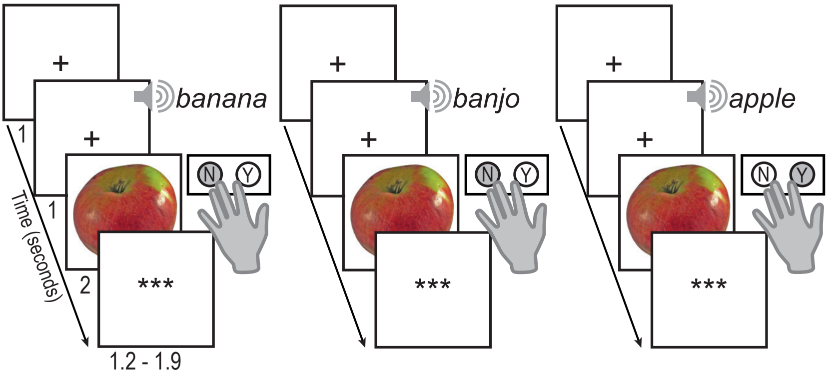


Figure 3: Trial structure. Examples are given for a related (left), unrelated (middle), and congruent (right) trials.

Analysis

The behavioural data and the analysis scripts are available on the OSF (see "Data Availability Statement" below).

Behavioral analysis

We analyzed the reaction times (RT) and the accuracies for the following three comparisons: related vs. unrelated, related vs. congruent, unrelated vs. congruent. The related vs. unrelated conditions are the main focus of our investigation as they provide a cleaner contrast of lexico-semantic competition, had the same number of trials per condition, and required the same 'no' response. The other two contrasts involving the congruent condition are reported for completeness. However, we note that these comparisons are more problematic. They involve different responses (and response buttons) and they had different number of trials (the congruent condition twice as many). Moreover, certain pictures with lower name agreement were responded to with 'no' in the congruent condition more often than expected. Finally, and most importantly, these contrasts do not isolate lexico-semantic competition as clearly and the congruent condition has converging information at various levels of representation (conceptual, lexical, phonological), making it very different from the other two conditions. Reaction times were analyzed via analysis of variance (ANOVA) with word-picture relatedness (related, unrelated and congruent) and subject type (patients and controls) as factors. Accuracy was analyzed via log-linear regression analysis with word-picture relatedness (related, unrelated and congruent) and subject type (patients and controls) as factors. The analyses were performed using JASP (JASP Team, 2018).

Hierarchical Drift Diffusion Model

It has been recently proposed that evidence accumulation is a plausible theoretical and empirical model of processes of lexico-semantic retrieval (Anders et al., 2017). We analyzed choice reaction time data with the hierarchical drift-diffusion model. The hierarchical Bayesian framework provides simultaneous estimation of individual and group subject parameters, which leads to robust model parameter estimates that better account for variability, especially in the performance of patients (Wiecki, Sofer, & Frank, 2013). Another advantage is the ability to obtain accurate model fits to the data using fewer trials, an extremely important aspect when investigating clinical populations. This makes HDDM an appropriate analytic tool for patient-based studies such as the current one (Ratcliff & Childers, 2015). We carried out the analysis using the Python toolbox HDDM 0.6.0 (Wiecki et al. 2013). We fitted models with different combinations of free parameters to the data via Markov Chain Monte Carlo (MCMC) fitting routines. We coded as "correct" a 'no' response in the related and unrelated conditions, and a 'yes' in the congruent condition. As commonly done using HDDM (Cavanagh et al., 2011; Zhang et al., 2016.), for each model we assessed a) the convergence rate of the numerical fitting routines and b) the ability of the model to capture the observed RT distributions. We excluded from further analysis the models that failed to reach convergence or failed to capture the observed RT distributions. Finally, we compared the models of interest by computing the relative Deviance Information Criterion (DIC), which provides a measure of the goodness of the model fit to the data while penalizing model complexity (Schwarz, 1978). Further, for the best fitting model, we examined the differences in posterior estimates by conducting a Bayesian ANOVA (since the model parameters are inherently not statistically independent) with type of participant and degree of semantic relatedness as factors. In the Bayesian framework, a Bayes factor (BF) provides graded evidence in favor or against the tested hypothesis (Wagenmakers et al., 2018). The evidence in favor of the alternative hypothesis (H_1) as opposed to the null hypothesis (H_0) has a subscript of 10 in the Bayes factor abbreviation BF_{10}, that is the evidence in favor of the alternative hypothesis. The following range from Kaas and Raftery (Kass & Raftery, 1995) facilitates interpretation of the Bayes factor: BF_{10} = 1-3 indicates “anecdotal” (“not worse more than a bare mention”) evidence for H_1 compared to H_0 ; BF_{10} = 3-20 indicates “positive” evidence for H_1 compared to $H_{0;}$ BF_{10} = 20-150 indicates “strong” evidence for H_1 compared to $H_{0;}$ BF_{10} >150 indicates “very strong” evidence for H_1 compared to H_0. We will only consider “strong” and “very strong” evidence for interpretation of our findings.

Results

Behavioral analysis

The results of RT and accuracy analyses are summarized in Figures 4 and 5. In all subjects, word-picture relatedness (related, unrelated, congruent) modulated the RTs (main effect: $F(2, 6583) = 238.26$, $p < .001$). Details of the statistical results for the post-hoc comparisons are presented in Table 3. Post-hoc comparisons revealed significant differences between the related and unrelated conditions (mean difference = 92.61). We also found that performance depended on the subject type (main effect: $F(2, 6583) = 26.62$, $p < .001$). We did not find an interaction between word-picture relatedness and subject type, $F(4, 6583) = 0.59$, $p = .66$. For the remaining post-hoc comparisons, we found a difference for related vs. congruent (mean difference = 196.33) and unrelated vs. congruent (mean difference = 103.73).

We performed a log-linear regression to test the effects of word-picture relatedness and subject type on accuracy. Accuracy was marginally not modulated by subject type (main effect of subject: Deviance (2, 6589) = 5.74, p = .057). Word-picture relatedness was not significant: Deviance (2, 6587) = 4.03, p = 0.13). We did not find any significant interaction between type of subject and semantic relatedness: Deviance (4, 6583) = 1.18, p = .88).

Table 3 Post-hoc comparisons related to the main effects of picture relatedness and subject type.

Contrast: Reaction times	SE	t	p
related vs. unrelated	10.6	8.72	< .001
related vs. congruent	9.2	21.32	< .001
unrelated vs. congruent	9.09	11.41	< .001
frontal vs control	9.23	5.34	< .001
temporal vs control	9.04	6.43	< .001
frontal vs temporal	10.62	0.82	0.68

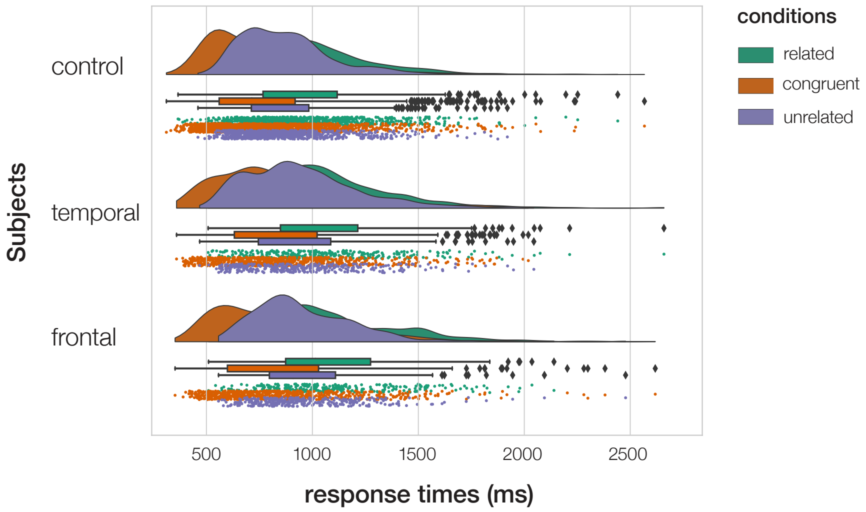


Figure 4: Single-trial response times for all subjects and experimental conditions.

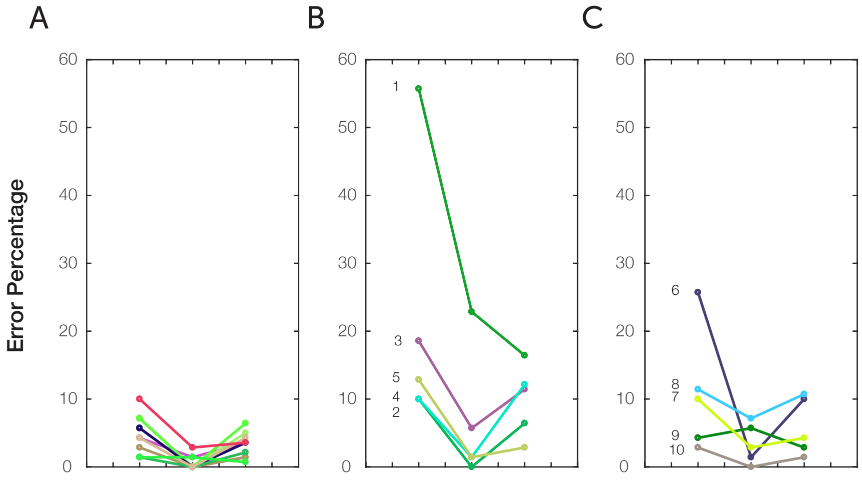


Figure 5: Individual percentage of errors for control subjects **(A)**, subjects with temporal lesions **(B)** and with frontal lesions **(C)**. R – related, U – unrelated, C – congruent conditions.

Hierarchical Diffusion Modeling

Model Convergence and Model Fit

We performed an outlier-removal on the data (fixed probability of 5 %) before feeding it to HDDM. For all analyses, we set the MCMC (Gelman & Rubin, 1992) fitting routines to 20,000 iterations with a burn-in period of 10,000 iterations and a thinning of 5. We assessed model convergence by examining the posterior sample (no autocorrelation found, no “spikes” in the posterior trace per each of the conditions, see Figure 6) and the R-hat statistic, which is a measure of convergence among multiple MCMC chains (three for the present study). Inspection of posterior density estimates revealed that the traces were stable over multiple samples, which indicated that the fitting routines have converged to a fixed estimate. The R-hat statistic under 1.1 indicated that chains with different starting values have converged to the same estimate. Successful convergence was confirmed also by a MCMC error for all of the parameters smaller than 0.01.

Finally, we evaluated the adequacy of each model’s predictions by examining the predicted posterior quantiles for the RT distributions for each participant (Figure 7).

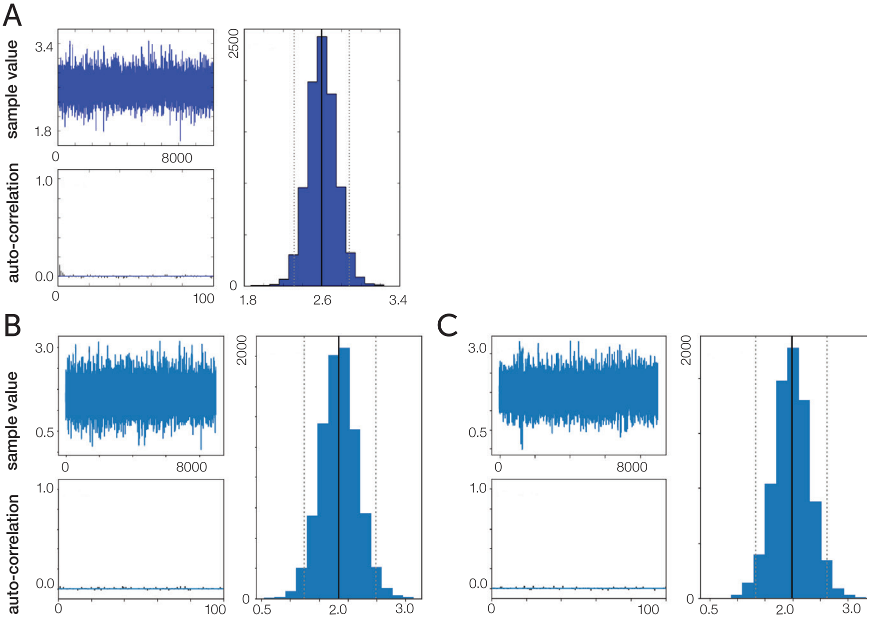


Figure 6: Example of a converged chain for control subjects **(A)** and patients with temporal lesions **(B)** and frontal lesions **(C)**. For each of the panels, top left: posterior trace, bottom left: autocorrelation, right: posterior histogram (solid black lines denote posterior mean and dotted black lines denote 2.5 and 97.5 percentiles).

Further, we quantitatively compared the models of interest by computing the associated DIC score for each model. According to DIC logic, a model with a lower DIC score is to be preferred to an alternative model with a higher DIC as the most parsimonious explanation of the data. Separately for patients and controls, we fitted seven variants of the model (see all model designs in Figure 8).

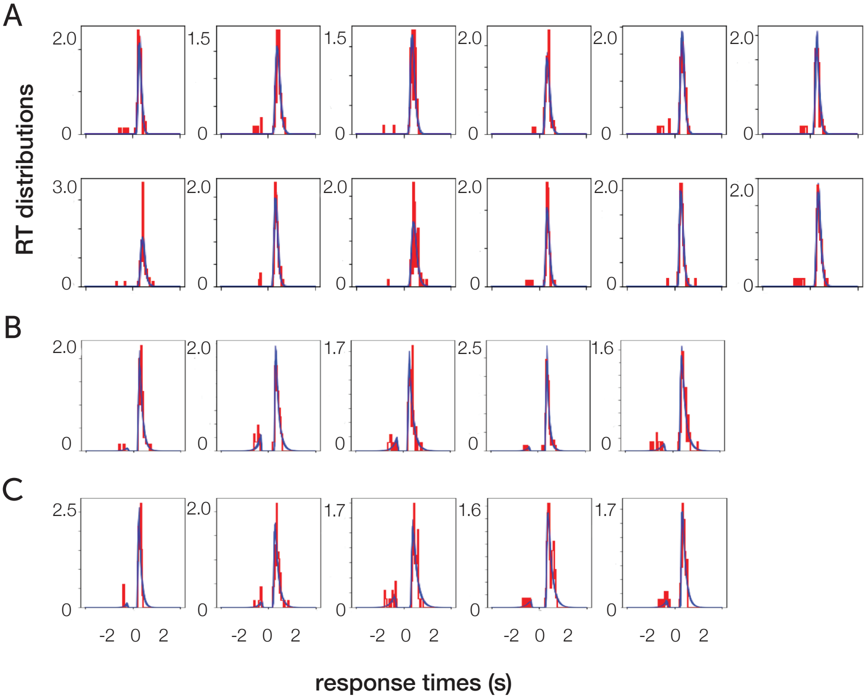


Figure 7: Observed RT distribution (red lines) and predicted posterior (blue line) from the model at hand (here, model-winner, i.e. model 1) for **(A)** control subjects, **(B)** subjects with temporal lesions and **(C)** frontal lesions. The related condition is shown in all panels. Negative RTs represent trials in which the response was incorrect.

First, we left drift rate, decision threshold, and non-decision time free to vary (Model 1, Figure 8) over the type of word-picture relatedness (related, unrelated, congruent). Then we fixed each of the parameters for the experimental factor of interest (semantic relatedness) across different models, leaving other conditions free to vary (Model 2-4, Figure 8). Further, we fixed two parameters leaving one to vary (Model 5-7, Figure 8). Next, we compared models with different combinations of parameters to find the most parsimonious account of the data. For both patients and controls, the model that best describes the data (i.e., the model with the lowest DIC score, see Table 4) was the model that allowed drift rate, threshold and non-decision time free to vary across the type of word-picture relatedness condition (Model 1 in Figure 8). Conventionally, a DIC difference of more than 10 indicates that the evidence in favor of the model-winner is substantial (Burnham, Anderson, & Burnham, 2002).

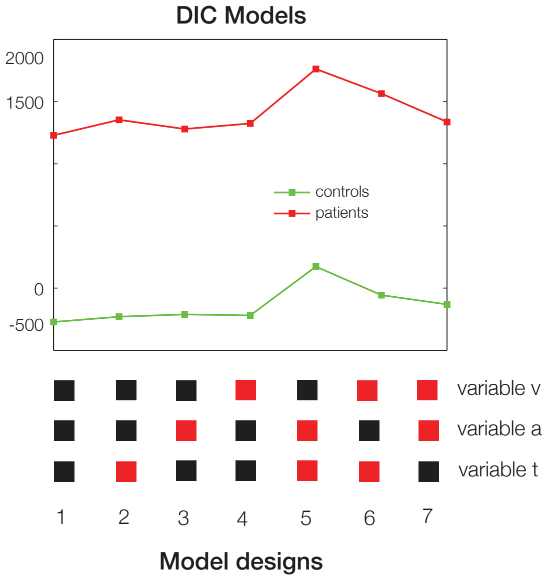


Figure 8: The deviance information criterion (DIC) value for each of the competing models. The models differ according to the number of parameters free to vary over the experimental factor semantic relatedness (related, unrelated, congruent). Black squares indicate that a parameter is free, red squares mark fixed parameters.

Table 4 Deviance information criterion for each model for patients and controls. For both controls and patients, the best winning model had variables drift rate, decision threshold and non-decision time free (Model 1, Figure 8).

subject type/model	**1**	**2**	**3**	**4**	**5**	**6**	**7**
controls	-272.78	-230.23	-211.23	-219.83	173.21	-56.36	-131.01
patients	1228.60	1353.65	1280.11	1324.27	1762.66	1564.54	1336.42

For the controls, the difference between the winning model (Model 1, DIC -272.78) and the second-best model (Model 2, DIC -230.23) exceeded the difference of 10. For the patient group, the difference between the first (Model 1, DIC = 1228.60) and second-best model (Model 3 = 1280.11) exceeded the difference of 10. We consider this sufficient evidence for postulating that Model 1 best describes the data. Therefore, we carried out further analyses focusing on this winning model's parameters.

Model Parameter Analysis

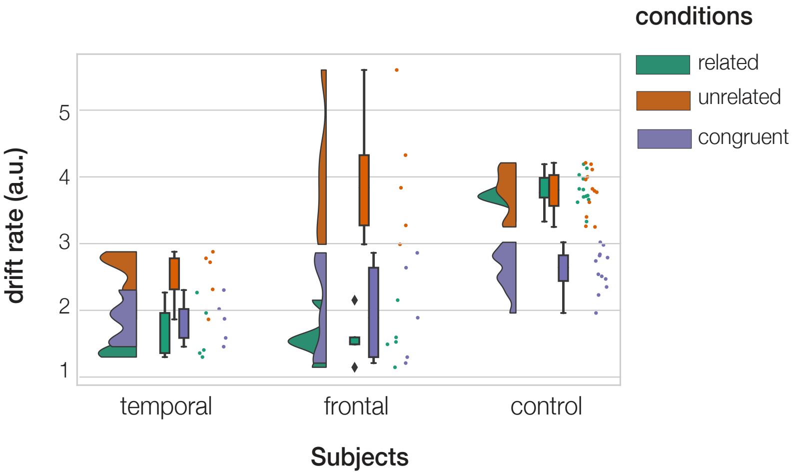


Figure 9. Posterior estimates of the hierarchical drift-diffusion model for the drift rate parameter. A difference in the related vs. unrelated condition was not detected in control participants, but was present in patients with both temporal and frontal lesions.

Figure 9 shows the drift rate estimates for each participant as a function of subject type and relatedness. A Jeffreys-Zellner-Siow (JZS, Liang, Paulo, Molina, Clyde, & Berger, 2008; Rouder, Morey, Speckman, & Province, 2012) Bayes factor ANOVA with default prior scales revealed that the best-fitting model consisted of main effects of word-picture relatedness, type of subject and an interaction between word-picture relatedness and type of subject. This model was preferred over the null model, BF_{10} = 6.80 x 10^{14} and over the second-best model that included main effects of word-picture relatedness and subject type, BF_{10} = 2.09 x 10^{9}.

Further, we performed Bayesian paired samples t-tests for each subject type (controls, temporal, frontal) separately, which revealed the following differences in regards to the comparison of interest. In controls, the related and unrelated pairs were likely to be processed at the same drift rate (related vs. unrelated: BF_{10} = 0.30). For patients with temporal lesions, there was “positive” evidence in support of increased drift rate in unrelated relative to related conditions (unrelated > related: BF_{10} = 11.32). For the patients with frontal lesions, we found “very strong” evidence in favor of increased drift rate in unrelated relative to related conditions (unrelated > related: BF_{10} = 48977).

For completeness, we report the result for the additional contrasts. In control participants, we found that related pairs were likely to have higher drift rate as opposed to congruent pairs (related > congruent: BF_{10} = 300221). Similarly, unrelated pairs were more likely to have increased drift rate in comparison to congruent pairs (unrelated > congruent: BF_{10} = 115.481). For patients with temporal lesions, we found "anecdotal" evidence in favor of a difference between related vs. congruent pairs (related vs. congruent: BF_{10} = 0.23) and "positive" evidence in favor of increased rate in unrelated in comparison to congruent conditions, (unrelated vs. congruent: BF_{10} = 7.83). For the patients with frontal lesions, we found "strong" evidence in favor of increased drift rate in unrelated in comparison to congruent conditions (unrelated > congruent: BF_{10} = 62.60), whereas related and congruent conditions showed "anecdotal" evidence supporting the condition differences (related vs. congruent: BF_{10} = 0.85). Additional analyses on selected groups of patients, reported in the Supplement, give further support to these findings.

In sum, we did not find reliable evidence supporting a difference in drift rate between related vs. unrelated contrast in control participants. By contrast, both temporal and frontal patients had increased drift rate for unrelated in comparison to related word-picture pairs.

Decision threshold

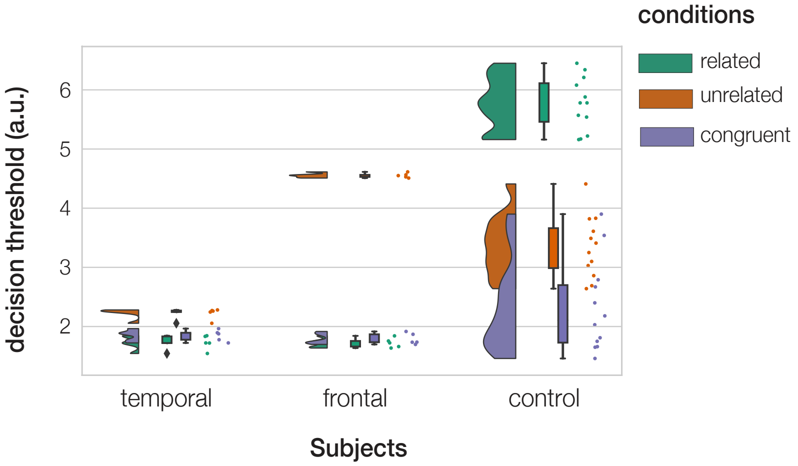


Figure 10: Posterior estimates of the hierarchical drift-diffusion model for the decision threshold parameter. Related vs. unrelated differences in decision threshold had the opposite directionality for control participants versus participants with temporal or frontal lesions.

Figure 10 shows the decision threshold estimates for each participant as a function of subject type and relatedness. A JZS Bayes factor ANOVA with default prior scales revealed that the model-winner was comprised of main effects of word-picture relatedness and type of subject, and their interaction. This model was preferred over the null model, BF_{10} = 6.68 x $10^{24}$and over the second-best model that included main effects only BF_{10} = 2.45 x 10^{6}.

In control subjects, we found “very strong” evidence that related pairs were more likely to have higher decision threshold as opposed to unrelated pairs (related > unrelated: BF_{10} = 1.08 x 10^{8}). For the patients with frontal lesions, there was “very strong” evidence in favor of higher decision threshold in unrelated vs. related pairs (unrelated > related pairs: BF_{10} = 48977). Thus, the direction of the effect was reversed in comparison to the controls. For patients with temporal lesions, we found “strong” evidence in favor of increased decision threshold in unrelated compared to related conditions (unrelated > related: BF_{10} = 43.06). Once again, the direction of the effect was reversed in comparison to controls.

For additional contrasts, we found “very strong” evidence that control subjects were more likely to have higher decision threshold for related in comparison to congruent condition (related > congruent: BF_{10} = 1.31 x 10^{7}). We also found “strong” evidence in favor of increased decision threshold for unrelated compared to congruent pairs (unrelated > congruent: BF_{10} = 96.37). For patients with frontal lesions, there was “very strong” evidence in favor of increased decision threshold for unrelated in comparison to congruent conditions (unrelated > congruent: BF_{10} = 68985.08). By contrast, the evidence supporting the difference between related and congruent conditions was “anecdotal”, (related vs. congruent: BF_{10} = 1.67). We observed a similar pattern for patients with temporal lesions. We found “anecdotal” evidence in favor of the difference between related vs. congruent pairs (related vs. congruent: BF_{10} = 0.80), but “very strong” evidence in favor of increased decision threshold in unrelated vs. congruent conditions (unrelated > congruent: BF_{10} = 307.52). Additional analyses on selected groups of patients, reported in the Supplement, give further support to these findings.

In sum, we found control subjects had a higher decision threshold for related as opposed to unrelated conditions. By contrast, the decision threshold was lower for related relative to unrelated word-picture pairs in both patient groups when compared to controls.

Non-decision component

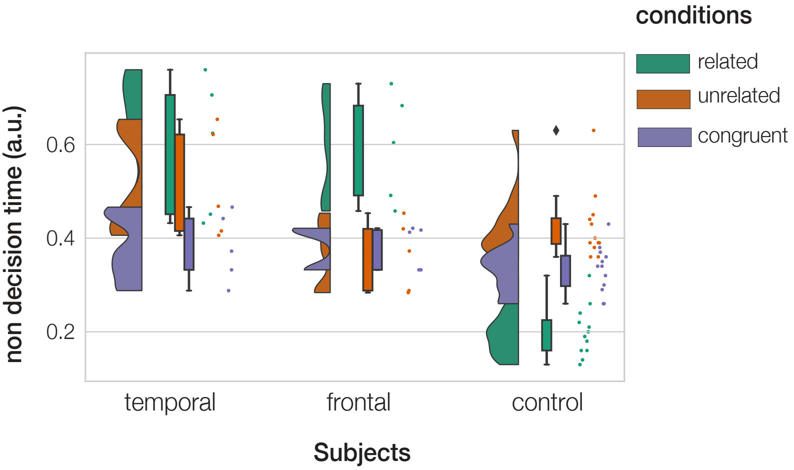


Figure 11: Posterior estimates of the hierarchical drift-diffusion model for the non-decision time parameter (Ter). Differences in non-decision time had the opposite direction in controls and patients with frontal lesions. Patients with temporal lesions did not show a reliable difference between related and unrelated conditions.

A JZS Bayes factor ANOVA with default prior scales revealed that the model-winner included main effects of word-picture relatedness, type of subject and their interaction. This model was preferred over the null model BF_{10} = 2.17 x 10^{10} and over the second-best model that included main effects of word-picture relatedness and subject type BF_{10} = 828.436.

We investigated the contrasts of semantic relatedness per subject group. In control subjects, there was "very strong" evidence in favor of decreased non-decision time in related as opposed to unrelated conditions (related < unrelated: BF_{10} = 4.39 x 10^{6}). By contrast, in subjects with frontal lesions, there was "moderate" evidence for increased non-decision time for related in comparison to unrelated pairs (related > unrelated: BF_{10} = 9.39). We found "moderate" evidence in favor of the difference between related vs. unrelated pairs (related vs. unrelated: BF_{10} = 6.57) in participants with temporal lesions.

For additional contrasts, in control participants, we found "very strong" evidence in favor of decreased non-decision time in related as opposed to congruent conditions (related > congruent: BF_{10} = 1912). By contrast, more non-decision time was needed for unrelated than congruent conditions (unrelated > congruent: BF_{10} = 31.38). In subjects with frontal lesions, we found an "anecdotal" difference between unrelated vs. congruent conditions (BF_{10} = 0.44). We also found that less non-decision time was needed for congruent in comparison to related items (congruent > related: BF_{10} = 28.01). Patients with temporal lesions needed less non-decision time for congruent in comparison to related (congruent > related: BF_{10} = 26.22) and for congruent compared to unrelated items (congruent > unrelated: BF_{10} = 22.27).

In sum, control participants needed less non-decision time for related than unrelated conditions. By contrast, frontal patients had the opposite pattern: increased non-decision time for related than unrelated conditions. Temporal patients did not exhibit meaningful differences between related and unrelated conditions.

Discussion

This study investigated the mechanisms that contribute to semantic control in healthy participants and participants with lesions in left temporal or frontal regions. Participants performed a word-picture matching task, in which they had to correctly verify whether congruent, semantically related and unrelated words matched pictures.

From the analysis of behavioral data, we found a semantic interference effect in RTs and errors for both patients and controls: related word-picture pairs resulted in longer RTs and more errors compared to the unrelated pairs. Patients on average had a higher error rate than control subjects, especially for the related condition. We note that the congruent condition is necessarily included in this task in order to give participants a task. However, we focused on the semantic interference effect, given that it better enables the study of semantic control. As we mentioned earlier, behavioral data alone does not allow exploring the nature - executive or/and semantic - of semantic interference. Therefore, we used a combination of RTs and accuracy to better understand latent cognitive processes underlying the process of semantic competition in the word-picture paradigm. We found that, for control participants, no reliable differences existed in drift rate between related versus unrelated word-picture pairs, whereas for both patient groups, an increased drift rate was found for unrelated in comparison to related word-picture pairs. Regarding the decision threshold, whereas controls had a higher decision threshold for related as opposed to unrelated conditions, the patients showed a lower decision threshold for related relative to unrelated word-picture pairs.

We formalized the process of lexico-semantic selection via a computational model of binary decision making, DDM, which provides an account of the process of lexico-semantic competition and resolution in terms of evidence accumulation (Anders, Ries, van Maanen, & Alario, 2015; Ratcliff, Thapar, Gomez, & McKoon, 2004). Previous studies related to the investigation of competitive lexico-semantic selection in both comprehension and production (Anders et al., 2017; Voss, Nagler, et al., 2013) have shown that the process of lexico-semantic selection can be sufficiently described with two parameter types – the rate of evidence accumulation (drift rate) and the amount of evidence accumulation (decision threshold). While the drift rate can reflect the process of spreading activation (Gomez et al., 2013; Kinoshita et al., 2017; Voss, Rothermund, et al., 2013), the decision threshold reflects broad control mechanisms (Anders et al., 2017). We found that, for healthy controls, task demands in word-picture matching tap more into control mechanisms, rather than into the process of spreading activation through the lexical-semantic system. We also found that patients with temporal and frontal lesions have similar cognitive profiles with respect to mechanisms of control and spreading activation described by the drift rate and decision threshold, as established by the model.

Control participants. The mechanisms of lexico-semantic competition are usually discussed in terms of lexical activation and lexical selection when there is the need to map semantic features to lexical items or vice versa (Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Levelt, Roelofs, & Meyer, 1999; Roelofs, 1992). While the process of activation has been quite extensively studied (Dell et al., 1997), the mechanisms of selection are still debated. According to many models of word production (Howard et al., 2006; Oppenheim et al., 2010; Roelofs, 2018), semantic features from several competing representations trigger the spreading activation in the semantic network, with the item with the highest activation level as a winner. It is plausible that the selection step requires executive control, especially in light of recent evidence dissociating activation and selection processes neurally (Piai, Roelofs, Jensen, Schoffelen, & Bonnefond, 2014). In the present study, we showed that in the context of lexico-semantic competition following the comprehension of spoken words, both high- and low competition conditions (related and unrelated word-target pairs) preserved the same level of activation in the semantic memory network (i.e., drift rate). Importantly, we showed that the selection of competing representations was supported by other mechanism potentially associated with executive control (i.e., decision threshold). This suggests that the mechanisms of semantic activation and selection may be dissociable (Nozari, 2017; Piai et al., 2014). Indeed, neuroimaging and lesion-based studies suggest that semantic selection (or control) and mechanisms of representation involve different brain areas (Jefferies, 2013; Lambon Ralph et al., 2017).

As for the “representation system”, it is generally assumed that the anterior temporal lobes are crucial for representing semantic knowledge (Patterson et al., 2007), and damage to these areas due to, for example, degeneration leads to the degradation of semantic representations (Guo et al., 2013; Lambon Ralph & Patterson, 2008). Among the brain areas associated with the control of semantic retrieval and selection are the left IFG and posterior MTG. Patients with damage in these areas show poor results in face of semantic distractors, retrieve irrelevant information for the task, and are helped by cues when retrieving information (Jefferies & Lambon Ralph, 2006; Jefferies et al., 2008; Thompson-Schill et al., 2002). In addition, a recent neuroimaging meta-analysis study showed that left IFG and posterior MTG are consistently activated in many tasks designed to tap semantic control (Noonan et al., 2013). In the present study, we do not consider the drift rate measures to necessarily reflect the “representation system”. In fact, none of our patients had semantic dementia, and some of the patients had both anterior and posterior parts of the left MTG damaged. Instead of indexing semantic knowledge per se, the drift rate may reflect the processes related to spreading activation in the lexico-semantic network. It is a common finding that the processes of lexico-semantic activation are tightly related to the left MTG (Baldo, Arévalo, Patterson, & Dronkers, 2013; Piai & Knight, 2018; Schwartz et al., 2009). By contrast, activity in frontal cortex is commonly interpreted as a control mechanism operating over lexico-semantic representations (Badre et al., 2005; Piai et al., 2014; Wagner, Paré-Blagoev, Clark, & Poldrack, 2001). However the contribution of left IFG in the resolution of lexico-semantic competition still remains elusive (de Zubicaray, Hansen, & McMahon, 2013; Piai & Knight, 2018; Piai et al., 2016; Piai, Roelofs, Acheson, & Takashima, 2013a; Python, Glize, & Laganaro, 2018). It is of importance to note that most of these studies address competitive word production. Studies investigating access to competing semantic representations using a word-picture matching task do not necessarily focus on the mechanisms of control per se but rather on the nature (lexical or semantic) of semantic interference (Campanella & Shallice, 2011; Harvey & Schnur, 2016). More broadly though, verbal comprehension mechanisms in both explicit (Demb et al., 1995; Spitzer et al., 1996) and implicit (Ruff, Blumstein, Myers, & Hutchison, 2008) semantic tasks involve left IFG, which is associated with control mechanisms of semantic retrieval. Whether the function of control is deployed by IFG, posterior MTG, or a combination of both and in which contexts remains an open question. Below we discuss the mechanisms of semantic interference in an instance of competition for a comprehension task (word-picture matching) in patients with lesions overlapping in left MTG vs. left inferior and middle frontal gyrus.

Frontal patients

Previously it has been shown that patients with damage to the left PFC, and particularly left IFG, fail to appropriately adjust their decision threshold, in order to handle the increased task difficulty with a higher degree of semantic interference (Anders et al., 2017). The authors argued that their finding supports the idea that left PFC, and specifically left IFG, subserve an interference control mechanism, potentially similar across different cognitive domains (Kan & Thompson-Schill, 2004; Ries et al., 2014). In the current study, we showed that frontal patients (with most lesion overlap in left IFG and left MFG) perform differently in an interfering condition from healthy participants. While healthy participants adjusted only the decision threshold but not the drift rate, these patients showed abnormal modulations in both decision threshold and drift rate. When presented with a "high demands" condition (semantically related), patients with frontal lesions were slower and made more mistakes (lower decision threshold and lower drift rate) in comparison to the "low demands" condition (semantically unrelated), where they were faster and more accurate (higher decision threshold, higher drift rate). In other words, these patients were not able to easily solve semantic competition (low drift and threshold). This is in line with previous studies showing that patients with damaged frontal areas after a stroke had problems with controlled semantic retrieval in picture naming, as patients' responses were driven mostly by irrelevant associations (Humphreys & Forde, 2005; Jefferies & Lambon Ralph, 2006). It seems that patients with frontal damage are better at the less competitive condition (unrelated) because of still functioning (albeit abnormally) excitation levels in semantic and control systems, which is not the case for the more difficult (i.e. related) condition.

The fact that both activation and selection processes are jointly impaired in the patients with lesions in frontal areas suggests that there is a certain degree of interaction between these processes (Nozari, 2017). For example, depending on the task, selection processes resolve competing representations either by adjusting the selection criteria dynamically from the activation levels of the lexico-semantic system or by adjusting the criterion post-hoc, after a certain level of activation in the lexico-semantic system has been reached (Nozari & Hepner, 2018). It may be that the overlap between activation and selection processes depends on the language modality (production vs comprehension), which is an important variable to account for when comparing lexico-semantic competition in word-picture interference paradigms.

Temporal patients

For temporal patients, we expected the manifestation of the deficit at the level of activation in the semantic network, leaving the selection processes critically impaired as a result of deficiency of spreading activation. We indeed found that patients with damage to the left MTG were critically impaired in both decision threshold and drift rate mechanisms, which we interpret as deteriorated levels of semantic network activation and executive components operating on those representations. Interestingly, the pattern for both drift rate and decision threshold followed the same direction (increased for unrelated in comparison to related conditions), in line with the findings of the frontal patients, which suggests an interaction between processes of semantic section and control.

Of note, the interpretation of the abovementioned results is highly dependent on the assumption that left posterior MTG contributes to lexico-semantic representations. However, the function of left posterior MTG remains somewhat elusive. Some studies propose that posterior MTG is a "knowledge hub" that links the associations between several concepts (Chao, Haxby, & Martin, 1999; Martin, 2007); other studies support the idea that posterior MTG participates in the control of semantic representations during retrieval (Davey et al., 2016; Noonan et al., 2013). A recent TMS study showed that left IFG and posterior MTG are crucial for semantic control rather than for semantic representation. Particularly, the study showed that these brain areas work in concert with each other, therefore the damage to either of these areas leads to disrupted manipulation of semantic knowledge (Whitney, Kirk, O'Sullivan, Lambon Ralph, & Jefferies, 2011). Similar executive malfunctioning in semantic retrieval tasks in patients with temporal and frontal lesions is in line with previous evidence revealing functional connections between IFG and temporal areas (Bourguignon, 2014; Grappe et al., 2018). Taking into consideration these recent results with regards to the contribution of the posterior MTG to the processes of retrieval of semantic information, we could conclude that posterior MTG contributes not only to the drift rate but also to the decision threshold.

Limitations

It is of course essential to pinpoint that lesions vary across the patients in our sample (even with the same focal damage), and the patients in our study have more than posterior MTG of PFC lesions, for example including the superior temporal gyrus (STG) and parietal regions.

Previously, it has been shown that temporo-parietal lesions can lead to deficits in semantic control (Jefferies & Lambon Ralph, 2006). Some patients in this study had lesions in three parietal brain regions: postcentral gyrus, supramarginal gyrus and inferior parietal gyrus. While to our knowledge there is no evidence that postcentral gyrus is involved in controlled semantic cognition, supramarginal gyrus and inferior parietal gyrus are thought to be part of the semantic cognition system. In the semantic control literature, parietal contributions (supramarginal gyrus and inferior parietal gyrus among others) are discussed in the context of temporo-parietal vs. prefrontal deficits. For example, persons with semantic aphasia, whose lesions are located not only in prefrontal cortex but also in temporo-parietal cortex manifested difficulties with controlling semantic representations (Jefferies & Lambon Ralph, 2006; Noonan et al., 2010, 2013). On the other hand, patients with inferior parietal or posterior temporal lesions did not exhibit impaired control over semantic activation in word production in comparison to patients with lesions in frontal cortex (Schnur et al., 2009). When it comes to the supramarginal gyrus, the evidence is not conclusive either. While there are studies showing the involvement of supramarginal gyrus in semantic control (Gennari, MacDonald, Postle, & Seidenberg, 2007; Nagel et al., 2008), a meta-analysis by Noonan et al. (2013) did not show significant activations in supramarginal gyrus, but rather within dorsal and anterior portions of angular gyrus, bordering the supramarginal gyrus.We inspected this issue in additional analyses after excluding patients with parietal lesions in frontal group to capitalize on the distinction between temporo-parietal vs. frontal lesions (reported in Supp. Table 2 and Supp. Figure 2). Experimental effects remained unchanged (i.e., higher decision threshold and drift rate for the unrelated in comparison to related condition), which makes an explanation in terms of parietal lesions alone driving the effects less likely. Altogether, given the somewhat unclear evidence with respect to the contribution of supramarginal gyrus and/or inferior parietal gyrus to semantic control, we interpret our results with caution regarding this issue.

It is important to note that two frontal patients have additional lesions in the STG. It has been suggested that the STG among other regions is associated with semantic representations rather than with control processes (Noonan et al., 2013). Thus, the PFC group might have shown additional disturbances in the lexico-semantic system (drift rate) in addition to the disturbed control exerted by the PFC (decision threshold). However, an additional analysis, which accounted for parietal contributions to the frontal group (see Supp. Table 2 and Supp. Figure 2), also excluded from the frontal group one patient with a large lesion in the STG. This additional analysis removed potential contributions of damage to the representation system from the frontal patients group. As reported above, experimental effects remained unchanged, which makes an explanation in terms of STG lesions contributing to our findings less likely.

Despite these control analyses, future studies, for example involving techniques such as TMS, are needed to explore the functional role of posterior MTG and left IFG in the context of degraded semantic retrieval with the modelling approach proposed here. Another limitation of the current study is that we cannot disentangle the contribution of the posterior and anterior portions of the MTG to the function of semantic control. Therefore, we cannot fully differentiate the contributions of control and representations to the process of semantic conflict resolution. Further studies should account for more precise differentiation of the lesions in order to clarify the function of semantic control in competitive selection in comprehension.

Conclusion

To conclude, we investigated the cognitive mechanisms of semantic interference in healthy subjects and patients with lesions overlapping in left MTG vs. left IFG/left MFG. We found that the driving force behind semantic interference in healthy subjects lies in the domain of executive control, rather than at the level of spreading activation in the lexico-semantic system. Patients with temporal and frontal lesions showed a similar pattern in the underlying mechanisms of semantic control in the context of word-picture matching, which suggests similar functional contributions of the posterior MTG and IFG to semantic competition during retrieval. These results support the notion that activation and selection are distinct but interactive processes that have to be preserved in order to resolve semantic interference efficiently.

6

Discussion

Words influence visual perception, something that has been suggested for at least half a century as the Sapir-Whorf hypothesis (Whorf & Carroll, 1956), which postulates that language affects the way we think. The mechanisms of this phenomenon remain a matter of debate to this day. Recent neuroscientific methods and theoretico-computational tools introduce new opportunities for investigating this long-standing debate. In this thesis, I aimed to shed light on the neural mechanisms and computations involved in the modulation of visual perception by lexico-semantic information introduced by words. I used a combination of neuronal data (fMRI, MEG, neuropsychology) and computational tools (DDM) in order to address the issue.

In this work, I capitalize on the decision making framework of DDM, through the lens of which I consider the experimental results. DDM is a biologically plausible computational model of decision making (Smith & Ratcliff, 2004), which suggests that the process of decision making can be formalized via stochastic accumulation of evidence towards a decision threshold. This allows us to make a parallel between visual perception and visual decision making in order to test to which extent lexico-semantic influences on visual perception can be explained by decision making mechanisms.

In **Chapter 2**, I showed that both associations ("tie") and identity words ("man") in the response priming paradigm led to a facilitation of face gender categorization (i.e., responding "man" or "woman" to faces). This effect was mapped to both speed of target processing and speed of motor response: when related to the target, both prime types resulted in increased drift rate and faster non-decision component. In **Chapter 3**, employing fMRI, I showed task-dependent modulations on the speed of visual categorization. Particularly, in the classification task (i.e., responding "man" or "woman" to faces), words congruent to the target (word: "man", target: male face) resulted in increased drift rate in comparison to incongruent ones (word: "man", target: female face). These results were similar to the results in the experiment in **Chapter 2** and followed well-established results from the literature (Voss, Rothermund, et al., 2013). In the verification task, where an explicit match was required between the prime and the target (i.e., respond "yes" if both prime and target match), female-related words resulted in an increased drift rate regardless of the gender of the target face, which suggests that the classification and the verification tasks require different performance strategies to accumulate decision about a visually presented category. In **Chapter 4**, employing MEG, I showed that words help to decide about ambiguous pictures by increasing the speed of visual categorization (i.e., responding "face" or "house" to an ambiguous picture that contained overlapping images of both a face and a house). Particularly, words congruent with the response resulted in increased drift rate in comparison to incongruent ones, which was similar to the findings in **Chapter 2** and **Chapter 3** (Classification task). In **Chapter 5**, using a lesion-symptom approach, I showed that individuals with brain damage in temporal or frontal regions have their visual-based decisions affected in similar ways, but fundamentally different from neurotypical controls. In this chapter, I used the word picture matching task, in which participants had to respond "yes" or "no" to a match between a word and a picture. While task performance in controls was explained by the amount of information needed to make a decision (decision threshold), individuals with frontal or temporal lesions exhibited similar dysfunctional modulation of both drift rate and decision threshold. These results support a diverse but interactive role of lexical-semantic memory and semantic control mechanisms.

Below, I discuss how the assumption from DDM that information accumulates as a random walk process can be instrumental in advancing the topic of language-vision interactions. Particularly, I will further discuss the importance of a theoretico-computational approach in conjunction with the neural data, its limitations and cautionary notes with regards to the topic of language and visual perception.

DDM in experimental neuroscience: modus operandi

In my thesis I illustrate three ways in which DDM might be useful for experimental cognitive neuroscience. First, I show that by using main constraints enforced by the model theory (i.e., it disentangles decisional vs. non-decisional processes, as discussed in Chapter 1) it is possible to disentangle cognitive vs. non-cognitive processes (**Chapter 2**). Second, DDM can be used to provide a mechanistic description of the process underlying neural effects (**Chapter 3 and 4**). Third, I use DDM to test whether it can formalize existing verbal theoretical models (here the model that describes the flexible use of language, **Chapter 5**). I will first outline how model constrains enforced by its theory are useful to study language influences on visual decision making. Then I will introduce the levels of informational description for a cognitive system (Marr's levels), followed by a discussion on how DDM can clarify the nature of neural representations. Finally, I will illustrate how the DDM machinery can be instrumental in formalizing the existing verbal theory of semantic cognition.

Using the constraints of DDM model to study how language influences visual decision making

In the field of language-perception interactions, signal detection theory (SDT, see **Chapter 1**) is used as a common measure of task performance, which describes how language has an effect on decisions. For example, it has been shown that language can boost the speed and sensitivity of visually driven decisions (Lupyan, Spivey, Frank, Wu, & Wade, 2010; Meteyard et al., 2007). The results however have not been consistent even in the studies with virtually the same experimental design. Here, I briefly consider two studies where words describing the direction of dot motion were followed by actually randomly moving dots (Francken et al., 2015; Meteyard et al., 2007). Behavioral findings in both studies indicated that participants are faster and more accurate about motion direction detection when word and visual motion are described coherently. However, these behavioral results translated quite differently to the measures of performance (sensitivity and bias, see BOX1 in **Chapter 1**). Particularly, words incongruent with subsequently presented visually random-dot-motion (RDM) resulted in impaired perceptual sensitivity in comparison to congruent word–RDM pairs (Meteyard et al., 2007). These results were interpreted as suggestive that words exhibit a top-down effect on visual perception at the level of stimulus-related sensory information. Another study, however, did not manage to find differences in sensitivity effects for congruent vs. incongruent word-RDM pairs (Francken et al., 2015). Instead they found that participants set a more liberal threshold for congruent as opposed to incongruent word-RDM stimuli. According to this view, language primes modulate decision criterion, which implies that stimulus distributions remain the same but the decision rule changes (see SDT, Macmillan & Creelman, 2005). Hence it is still debatable how behavioral performance is affected by language, and whether language modulates perceptual representations of visual stimuli.

Besides the differences in experimental procedure (e.g., Meteyard et al.'s study was held on two different days, which might have led to habituation effects), it is important to consider an intrinsic feature of SDT that is crucial for the interpretation of language supported visual decisions. In SDT, the decision making strategy (sensitivity or threshold) is reflective of various processes such as motor execution / preparation and perceptual encoding, in addition to actual the process of decision formation (see red frame on Figure 4, **Chapter 1**). For example, the fact that one has to press the button or reply "up"/"down" in the context of the abovementioned RDM experiment could lead to motor effects rather than to decisional effects but STD itself will not be able to differentiate them. The drift diffusion model instead provides a measure of "pure" decision variable formation, without additions of non-decisional components such as encoding or motor execution. In the case of response priming (**Chapter 2**), where the prime can affect both motor preparation time and the speed of decision formation (motor vs. cognitive effects), it was important to disentangle these process. I indeed showed that the prime does not bypass the cognitive evaluation of the target picture, leading to exclusively motor facilitation effects. Conversely, I showed that words promoted the speed of visual decisions in addition to having effects on non-decisional processes.

I will further elaborate on the meaning of the notion “the speed of visual decisions” promoted by language from the cognitive neuroscience perspective. Particularly, I will layout how a theoretical model (DDM) that describes how neurons collect information can be useful for better understanding higher cognitive functions, such as semantic cognition. I will further consider the use of the model in a hierarchy of knowledge description (i.e., Marr’s levels of analysis) within which we can describe a cognitive phenomenon. Finally, I will discuss the results of this thesis within the levels of the model.

Algorithmic level of information description as a bridge between brain computations and cognitive theories

According to Marr (Marr, 1982), systems that process information can be analyzed at three different levels depending on what information is computed at each of them. While the *implementational* level is concerned with the physiological structure that supports the behavior, the *algorithmic* level concerns representations and processes that drive behavior. Finally, the *computational* level, the most abstract of Marr’s levels, is concerned with the “abstract model description” (see **Figure 1 A**). Previously (Love, 2015), a computer metaphor was used to represent these levels, which consists in the following. For a computer, the *computational* level is a program description, the *algorithmic* level is the program itself and the *implementation* level is the hardware. Taken together, Marr’s levels provide a holistic perspective on what, how and why a certain cognitive phenomenon functions. While being generally accepted, in the past decades, due to reductionalistic tendencies in the field of neuroscience, the use of this model was put in doubt (Peebles & Cooper, 2015). Particularly, the algorithmic level of the model was downplayed by neuroimaging approaches that do not quite distinguish between *algorithmic* and *implementational* levels. The common criticism of the *algorithmic* level referred to the fact that multiple algorithms (i.e., functions) can create the same output, similarly to how computer programs can accomplish the same task. The algorithmic level can be referred to as the famous “black box” (see **Figure 1 B**),

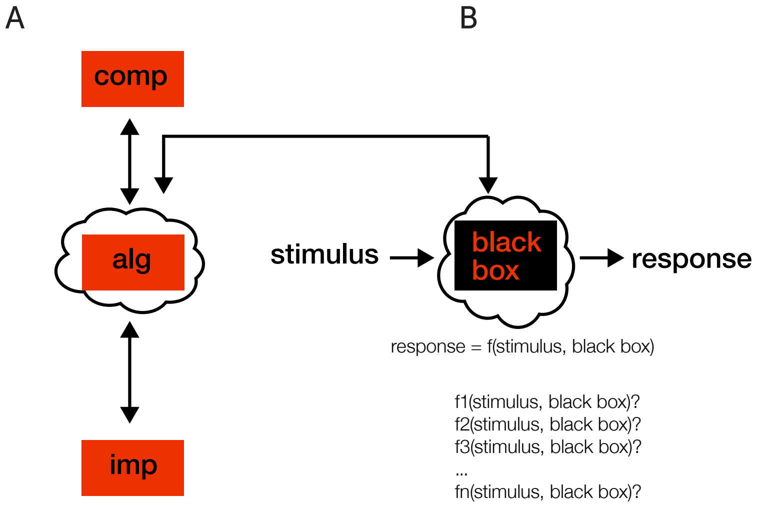


Figure 1: (A) Marr's levels of representation: IMP = implementational level; ALG = algorithmic level; COMP = computational level. **(B)** The "black box" of psychics.

i.e., the metaphor for non-directly observed computations of our brain that subserve certain cognitive function. Deciphering the "black box" is arguably the goal or neuroscience.

DDM is a biologically plausible cognitive model that attempts to explain cognitive mechanisms. DDM can be used as a tool to validate and / or develop a cognitive theory by offering an intermediate explanatory level between brain computations and cognitive abstractions. This feature makes DDM an *algorithmic* level model. Moreover, *algorithmic* models can work bi-directionally. One the one hand, they help understanding brain computations that support cognition. On the other hand, neuroimaging (brain data) evaluates mechanistic accounts of how cognitive processes unfold in the brain and ultimately helps to select among competing algorithmic models.

Eventually the more we learn about certain cognitive function, cognitive models may be updated with new features, or even new models that reflect better the neuro-cognitive architecture can emerge. The use of DDM as an algorithmic tool provides a dynamic interaction between implementational and computational levels resulting in a scientific framework that is promising for understanding how certain cognitive phenomena function.

To sum up, DDM as a theory at the algorithmic level of computation provides a bridge between cognitive theory and actual brain computations. I will further discuss the results in this thesis using Marr's levels of analysis, which provides us with different levels of scientific abstraction.

Language-perception interaction through the lens of DDM

It has been proposed that language influences visual perception via modulation of the activity in the brain regions at the low level of hierarchy that is associated with stimulus processing (Amado et al., 2018; Landau et al., 2010). While tempting, the claim that language affects early stages of perception can have an alternative explanation. It can be that activity in visual cortex represents feedback loops from the other areas rather than coding low-level visual features. Many of the studies investigating language-vision relationships are indeed fMRI studies, which is a method known for its sluggish temporal resolution. Therefore fMRI is limited in resolving brain activity at the level of milliseconds, which is a natural pace for language. This opens up the possibility for the alternative explanation i.e., the effects detected in visual areas might reflect the feed-back loops occurring at a later timescale. One way to disentangle this is to use another neuroimaging modality, such as MEG or EEG (the approach adopted in **Chapter 4**), which are known for their more refined temporal resolution. Another approach would be to turn to a combination of computational methods and neuroimaging tools that could shed light on the nature of language-vision interactions.

As I mentioned earlier, DDM is a computational model that describes the process of decision making via a set of parameters that guide human performance. It is an algorithmic-level model, which provides a bridge between different levels of computations in the brain. I used this feature of DDM in order to clarify the computational mechanisms behind the low-level activity introduced by the primes (**Chapter 3**). Specifically, I hypothesized that due to the feedback loops, the activity in sensory cortex can reflect the relatively late feed-back mechanisms formalized within the decision making approach. Particularly, I tested whether the BOLD signal in sensory regions could be explained by the speed of decision variable formation, i.e., the speed of visual categorization.

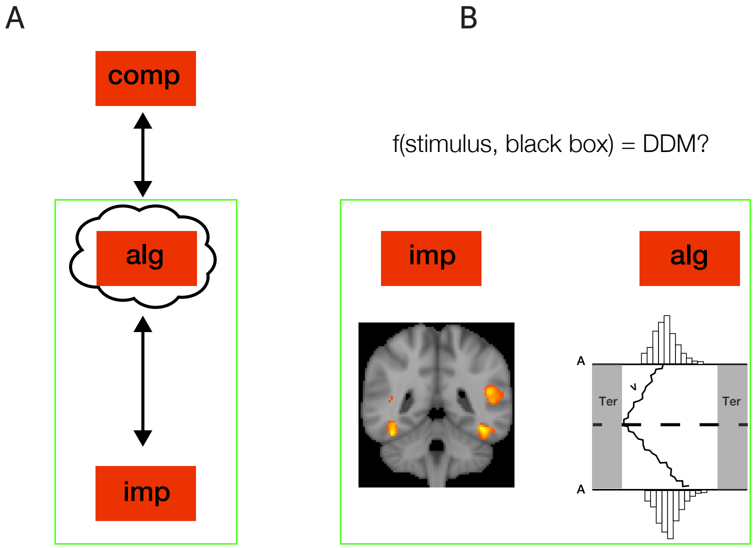


Figure 2: **(A)** Marr's levels of representation: IMP = implementational level; ALG = algorithmic level; COMP = computational level. **(B)** The hypothesis: the BOLD signal in sensory related brain regions reflects the computations behind the accumulation of decision variable (**Chapter 3**).

The results of our experiment (**Chapter 3**) did not show conclusive results. First, I did not find significant results when analyzing BOLD data related to language effects without model parameters. I argue that this was mostly due to experimental set-up limitations. The fact that the experiment consisted of two tasks, with few trials per condition, might have contributed to the low power of the study. When performing correlational analysis with the drift rate, I did not find results that would support my initial hypothesis either. It may be that these are reminiscences of the low power issues. To reiterate, the current results are not sufficient to either support or reject the idea that the BOLD signal in visual brain areas can be accounted for by the drift rate. . Future studies should investigate this issue further.

In **Chapter 4**, I further investigated whether markers of visual processing (M170 event-related component) can be influenced by words and further accounted for by DDM parameters. Instead of using fMRI (I adopted this approach in **Chapter 3**) with its intrinsically slow temporal resolution, I used another neuroimaging modality, MEG, notorious for its high temporal resolution. In addition to testing whether primes tap into visual processing (M170), I investigated whether words can affect

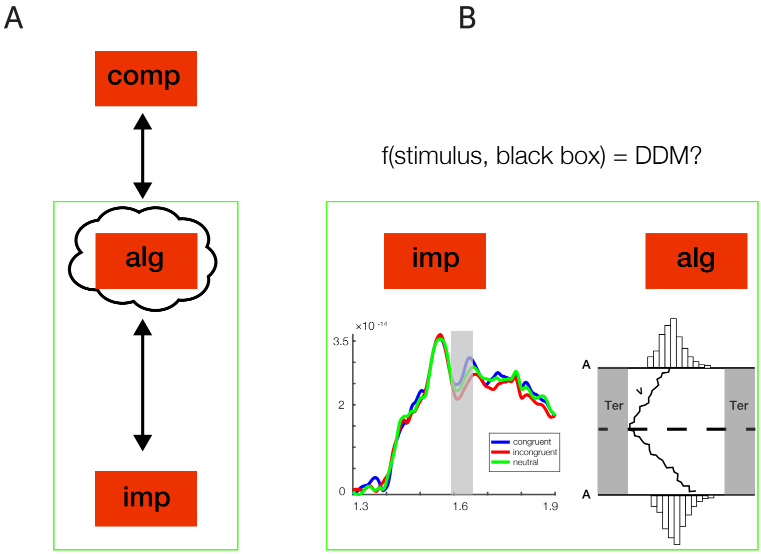


Figure 3: (A) Marr's levels of representation: IMP = implementational level; ALG = algorithmic level; COMP = computational level. **(B)** The hypothesis: markers of visual (M170) or semantic (N400m) processing reflect the computations behind the accumulation of decision variable (Chapter 4).

visual perception via semantic memory (N400m event-related component). I further investigated whether these markers were to reflect the speed of visual categorization, as reflected in the drift rate.

Unexpectedly, in this experiment, I did not find either N400m or M170 effects, but rather a P300 effect. This suggested that words help to recognize ambiguous input at the post-perceptual stage rather than at the stage of early visual perception. Since the P300 has been associated with many cognitive mechanisms such as attention and memory updating (Comerchero & Polich, 1999; Polich, 2007), decision confidence (Hillyard, Squires, Bauer, & Lindsay, 1971), and surprise (Mars et al., 2008), to name a few, its function is still unclear. Research in the field of semantic integration has shown that the P300 can be correlated with aspects of semantic access (Dorjee, Devenney, & Thierry, 2010; Hill, Ott, & Weisbrod, 2005; Hill, Strube, Roesch-Ely, & Weisbrod, 2002; Rossell, Price, & Nobre, 2003). Particularly, in the context of semantic memory tasks, the decrease in P300 amplitude has been associated with increased memory load (Mecklinger et al., 1992) . Thus, it could be that a decreased P300 reflects enhanced memory load for incongruent in comparison to congruent word-picture pairs. It has also been shown that the P300 may reflect the decision making process in the context of classification of degraded images (i.e., participants had to discriminate between a face and a house embedded in visual noise, Philiastides & Sajda, 2006). Therefore, we aimed at investigating the computational mechanisms of decision making by correlating P300 with drift rate.

Interestingly however, we did not find any correlations between P300 and drift rate as it was suggestive from the study previously investigating ambiguous object perception (Philiastides & Sajda, 2006). There might have been several reasons that could explain the absence of correlation results. First, I used the mean amplitude measure instead of the peak amplitude measure that was reported in the study that did find a correlation between P300 and behavioral measures (Philiastides & Sajda, 2006). Second, I performed correlations at the group level, not on a trial-by-trial basis. Third, it can be that inter-trial variability in evidence accumulation (drift rate) plays a critical role in the process of integration of language and vision information. For example, neural activity before the language prime can also affect the decision and response about the ambiguous target. To account for this, trial-by-trial variability needs to be added to the model. However in the present study this was not possible due to the larger number of trials required for such an analysis. Finally, it can be that the P300 effect in this task, rather than reflecting aspects related to decision making, reflects other mechanisms such as, for example, decision confidence (Hillyard, Squires, Bauer, & Lindsay, 1971; Squires, Hillyard, & Lindsay, 1973).Taken together, these analysis choices could have contributed to null correlational results. Further studies should be performed in order to clarify the issue.

Semantic cognition through the lens of DDM

In **Chapter 5**, I illustrated how DDM can formalize existing cognitive theory describing flexible language use, i.e., controlled semantic cognition. The theory of semantic control is a descriptive theory that postulates that language use, including tasks such as semantic categorization and matching, is subserved by the systems of semantic representation and control. In this thesis, I proposed that two cognitive processes described by the parameters of DDM – the speed of evidence accumulation and the decision threshold – can serve as proxy for the systems of semantic representation and control (i.e., broadly executive mechanisms). From this point of view, mechanisms of executive control (decision threshold) can be applied alongside with the system of semantic representation (drift rate), which would assure the flexibility of semantic cognition.

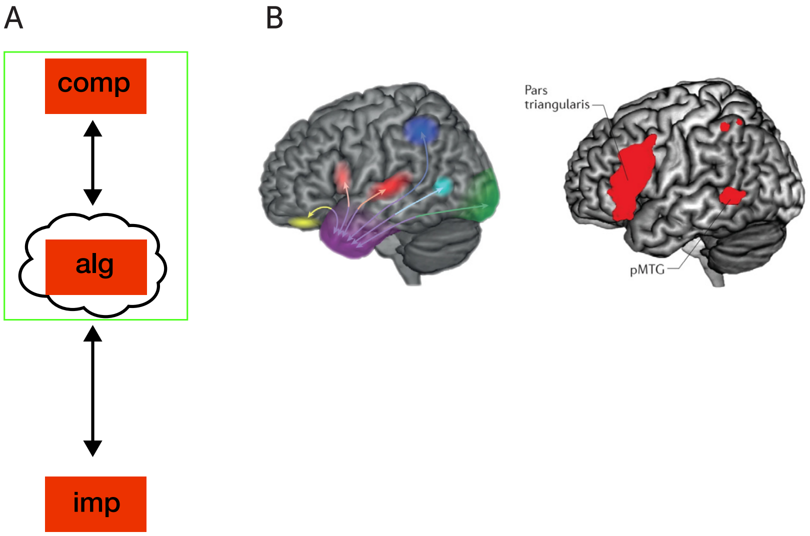


Figure 4: (A) Marr's levels of representation: IMP = implementational level; ALG = algorithmic level; COMP = computational level. **(B)** DDM as a tool that provides a link between cognitive theories and computations at the algorithmic level. Figure 4B is modified from (Lambon Ralph, Jefferies, Patterson, & Rogers, 2017).

This approach resulted in a few insights. First, experimental results on healthy controls suggested that the task effects were reflected in the decision threshold setting rather in the speed of evidence accumulation. I interpreted it as suggestive of executive rather semantic mechanisms needed to solve the word-picture matching task. In the absence of neuroimaging tools, this dissociation was only possible in the context of the DDM model. Second, lesion data showed that patients with frontal lesions showed dysfunctional drift rate and decision threshold for unrelated pairs relative to related ones, which was similar for the patients with temporal lesions. The computations required by word-picture matching were affected by left-frontal and left-temporal damage similarly, resulting in systematic deficits across lexical-semantic memory and executive functions. Ultimately, I illustrated that lexical-semantic memory and semantic control are diverse but interactive mechanisms essential for flexible language use. Besides having theoretical value, these results can contribute to language assessments strategies, highlighting that disruptive behavior in tasks typically used for diagnosis of language disturbances, such as word-picture matching, might be due to executive disturbances rather than semantic ones in patients with lesions in areas relevant for language.

Cautionary notes about DDM as a tool for modeling cognition

While drift rate represents the property of the neurons, its cognitive meaning might vary in the context of a certain experiment. For example, one can hypothesize that increased drift rate for congruent vs. incongruent prime-target pairs would reflect the modulation of attention over certain representations (semantic or visual) (see **Chapter 2**). It has been shown that attention directed to a particular stimulus enhances task-specific features of this stimulus, resulting in increased BOLD for relevant vs. irrelevant features (Brouwer & Heeger, 2013; Jehee et al., 2011). If the drift rate were to correlate (positively) with the BOLD in congruent vs. incongruent pairs, then the drift rate would be associated with the processes of attention deployment. However, in the case of a negative correlation between BOLD and the drift rate (increased drift rate for congruent vs. incongruent would correlate with decreased BOLD for congruent vs. incongruent) it can be that instead of attentional modulations, the drift rate reflects sensory habituation/adaptation that leads to efficiency of sensory processing. To sum up, caution is needed in order to interpret the cognitive meaning of the drift rate in the context of specific tasks.

Another point of discussion with regards to cognitive interpretations of DDM is whether drift rate reflects the processes of facilitation or interference. It is well known that increased drift rate indicates faster speed of processing. For example, congruent prime-target pairs result in increased drift rate in comparison to incongruent ones (**Chapters 2, 3**). However, in **Chapter 5**, patients with frontal or temporal lesions exhibit increased drift rate for unrelated vs. related items. Rather than reflecting the ease of processing, this finding reflects the compensatory performance of patients in order to solve the task. These results highlight that increased drift rate does not always reflect an advantage in processing (i.e., facilitation), but it has to be understood within the context of an experimental paradigm and participants characteristics. In conclusion, investigating requires making connections at different levels of informational processing. In this thesis, I showed how a model-based computational neuroscience approach can be useful in investigating how language affects visual decision making.

Limitations of the DDM approach

The major limitation of HDDM is an underlying assumption of how the brain works. It is assumed that the "brain extracts, per time unit, a constant piece of evidence from the stimulus (drift) which is disturbed by noise (diffusion) and subsequently accumulates over time" (Bitzer, Park, Blankenburg, & Kiebel, 2014, p.1). This framework differs from the Bayesian approaches, which view perceptual decision making as a probabilistic inference, in which evidence is used to update the probability of a current choice. Particularly, Bayesian models incorporate the perceptual uncertainty of the observer as a key parameter that would explain behavior. It has been recently proposed that neural activity in visual cortex measured with fMRI during simple perceptual decision making task carries a representation of the uncertainty in the visual input (van Bergen, Ma, Pratte, & Jehee, 2015). In spite of the seemingly major differences between the models, it has been recently shown that one can cast the drift and diffusion parameters of the Drift diffusion model as internal uncertainties of a decision making observer about sensory input (Bitzer et al., 2014). This is an active area of research, which would be interesting to validate with empirical studies, particularly in the field of language-vision integration.

Another limitation is related to the specifics of the model, which posits that effects on perceptual encoding cannot be clearly isolated from the effects on motor execution. Therefore, from a theoretical perspective one cannot draw any inferences about the absence or presence of language-vision integration effects at the early stage of visual perception since the model is not designed to capture those. However it is still an empirical question whether integration at the later time-scales (i.e. decisional) can be reflected in the regions associated with conceptual or semantic processing. I have addressed this question in Chapter 3 and Chapter 4 of this thesis.

Yet another potential limitation is routed in the fact that any binary decision making, regardless of the nature of the stimuli, can be represented via the process of accumulation of the evidence towards a decision threshold. This raises questions whether DDM can be informative about language specific processes in the context of language-vision integration. While I expand more on this issue in the following paragraph, where I discuss design constraints, language specific vs. domain specific evidence accumulation, here I point out that the decisional effects of language on visual perception in the context of two-alternative forced choice tasks are not conclusive either. For example, it has been found that words incongruent with upcoming random-dot-motion panels (rdm) resulted in impaired perceptual sensitivity in comparison to congruent word–random dot motion pairs (Meteyard et al., 2007). Yet another experiment with a similar design (Francken et al., 2015) did not manage to find differences in sensitivity effects for congruent vs. incongruent word-rdm picture pairs. Instead, they found that participants set more a liberal threshold for congruent as opposed to incongruent word-picture words. This discrepancy makes language-based decision making mechanisms worthy of further investigation.

DDM for language processes in the context of two-alternative forced choice tasks and beyond

In this thesis, I considered the process of language-vision integration via the lens of evidence accumulation. In order to describe evidence accumulation, two boundaries are necessary, which signify a binary choice (e.g., "yes" for the upper boundary and "no" for the lower one; "woman" for the upper boundary and "man" for the lower one, etc.). The model assumes that decisions are based on evidence accumulation, and the response is made once the evidence for a particular response reaches a threshold. The use of DDM in current language research can be two-fold: on the one hand one could take the path of disentangling purely linguistic processes vs. decision-related processes; on the other hand, one could investigate whether language processes can be explained by language-specific vs. domain-general evidence accumulation.

To illustrate the first point, many neuroimaging experiments, which investigate how we extract and use word meaning, have binary tasks (such as lexical decision, related/unrelated judgment between words, matching a word to a picture), which abstract away from our use of language in real situations. However, given decades of history of experiments with binary tasks in language research, revisiting those results with a DDM approach might be useful to shed light on some long-standing questions. For example, neural areas that showed modulation of indirect i.e. measured with blood-oxygen-level dependent signal (BOLD) neural activity in the context of semantic relatedness in binary tasks were ATL (Rossell et al., 2003), posterior temporal structures (STG) (Rissman et al., 2003; Ruff et al., 2008), frontal structures such as middle frontal gyrus (MFG) (Rissman et al., 2003) and inferior frontal gyrus (IFG) (Ruff et al., 2008) and precentral/postcentral gyri (Rissman et al., 2003; Ruff et al., 2008). These activations however can be modulated by the depth of semantic processing (e.g., a lexical decision task or a task about semantic relatedness), selection demands (focus oon global or specific aspects of the stimuli). The decision process related to the binary task is likely to be reflected in the neural results as well. Therefore one could use a DDM to identify brain regions that reflect the process of evidence accumulation, related to the binary decision making. This would help to critically access the neural underpinnings of semantic cognition. Furthermore, it is useful to demonstrate whether neural substrates behind evidence accumulation in two-alternative forced choice tasks (2AFC) vary for linguistic vs. non-linguistic stimuli. If no differences are found, then the argument towards the non-specificity of evidence accumulation would be well supported. Therefore all potential brain areas that would exhibit a correlation with the drift rate would reflect a non-language specific decision making process. However, if there is evidence towards language-specific evidence accumulation, it introduces new interpretation of neural computations of semantic cognition.

Drift diffusion models describe the process of decision making in two alternative force choice tasks (2AFC tasks), which are not necessarily present when we process language daily. To what extent is DDM informative about the integration of language and visual information? I believe that decision making can inform us about categorization, and ultimately provide a framework for understanding language influences on categorization. When I pick a ripe cherry or buy emerald color to paint the walls in the living room, esthetical/taste preferences aside, I rely on categorization. In order to categorize, I need to identify a set of candidate categories. Each of the candidate categories has to be evaluated and one category has to be selected. Formally speaking, there are other formal frameworks conceptually similar to DDM that can consider a task with more response choices (such as a linear ballistic accumulator model, LBA, (Donkin, Brown, Heathcote, & Wagenmakers, 2011)). The principle of choice selection is governed by the evidence accumulation for each choice up until the response threshold. Having the computational machinery which is able to capture the decision dynamics for multiple choices might help to capture category selection for more naturalistic tasks such as recognition. One could then further consider designing an experiment with a multiple choice object recognition embedded in a language context in order to test the effects of language on visual recognition. A combination of clever experimental design and a linear ballistic accumulation model would be a step towards a more ecologically valid approach of exploring language effects on categorization.

It is important to acknowledge that DDM/LBA approach is limited to the tasks that have a finite number of alternatives included explicitly in the task, therefore experiments with open answers or no answers at all (such as passive viewing settings) would not benefit from this framework. However the DDM perspective advances the field by raising the following questions: What are the computational differences between retrieval of semantic knowledge vs. manipulation of semantic knowledge? Are there multiple forms of control when manipulating/retrieving semantic knowledge? Are perceptual features of the objects accessed during manipulation of semantic knowledge as opposed to retrieval of semantic knowledge? Future studies are needed to address these questions.

In conclusion, in this thesis I formalized visuo-semantic interactions from a decision making perspective, i.e., as a process of evidence accumulation to the decision threshold. Particularly, I showed that the language advantage in visually driven decision making translates into the speed of evidence accumulation. Furthermore, I investigated whether neural markers underlying linguistically boosted visual decisions can be accounted for by the speed of evidence accumulation or by any of the others DDM parameters (i.e., decision threshold, starting point). In spite of the fact that I did not find correlations between neural and modelling data, I outlined directions for further analyses, which make this formal approach worth of pursuing further. Finally, I proposed the evidence accumulation framework as an exploratory platform for disentangling the processes of semantic control and interference. In sum, I investigated visuo-semantic interactions at different levels of scientific abstraction using a combination of behavioural, neuroimaging and formal methods, which makes the results and narrative of this thesis informative for future studies.

Supplementary material

SUPPLEMENT FOR CHAPTER 2

Table 1.1. List of associative words related to male and female concepts and neutral words.

Words related to male concept	Words related to female concept	Neutral words
vader (father)	zus (sister)	dag (dag)
baard (beard)	dochter (daughter)	adrenaline (adrenaline)
jongen (young man)	tante (aunt)	appelsap (apple juice)
snor (moustache)	mama (mother)	telefoon (phone)
kostuum (costume)	echtgenote (wife)	friten (fries)
hoed (hat)	rok (skirt)	loterij (lottery)
broer (father)	taille (blouse)	ticket (ticket)
zwaard (sword)	nicht (niece)	tunnel (tunnel)
kerel (guy)	parfum (perfume)	verlof (vacation)
farao (pharaoh)	dame (lady)	gazon (lawn)
stropdas (tie)	handtas (handbag)	koord(string)
bruidegom (groom)	hoofddoek (headscarf)	rivier (river)
geweer (weapon)	mascara (mascara)	vuur (fire)
vlinderdas (bow tie)	pruik (wig)	roem (fame)
sik (beard)	rouge (rouge)	maag (stomach)
zoon (son)	jurk (dress)	bos (forest)
pijp (pipe)	meisje (girl)	pen (pen)
smoking (tuxedo)	haarstukje (hairpiece)	kiosk (kiosk)
kilt (kilt)	panty (tricot)	bioscoop (cinema)
scheermes (razor)	mantelpak (suit)	rekening (bill)

Table 1.2. Similarity measures for associative words vs. labels in terms of cosine similarity (1 – cosine distance), where cosine distance is derived from word2vec model. Words with lower similarity score are considered to be more similar (man-man, 0) in comparison to words with higher similarity score (beard-man, 0.67).

Prime count (associations before division)	prime	target	Cosine similarity	Prime type (division of associative)
1	vader	man	0.471586	Label
2	baard	man	0.67291	Association
3	jongen	man	0.347639	Label
4	snor	man	0.669841	Association
5	kostuum	man	0.763772	Association
6	hoed	man	0.696146	Association
7	broer	man	0.465685	Label

Table 1.2. Continued.

Prime count (associations before division)	prime	target	Cosine similarity	Prime type (division of associative)
8	zwaard	man	0.661439	Association
9	kerel	man	0.274264	Label
10	farao	man	0.802513	Label
11	stropdas	man	0.709387	association
12	bruidegom	man	0.7429	Label
13	geweer	man	0.610479	Association
14	vlinderdas	man	0.762285	Association
15	sik	man	0.754113	Association
16	zoon	man	0.45961	Label
17	pijp	man	0.734082	Association
18	smoking	man	0.720216	Association
19	kilt	man	0.677306	association
20	scheermes	man	0.659975	association
1	zus	vrouw	0.472862	Label
2	dochter	vrouw	0.311131	Label
3	tante	vrouw	0.537335	Label
4	mama	vrouw	0.640638	Label
5	echtgenote	vrouw	0.363388	Label
6	rok	vrouw	0.750863	Association
7	taille	vrouw	0.800168	Association
8	nicht	vrouw	0.647979	Label
9	parfum	vrouw	0.690007	Association
10	dame	vrouw	0.452672	Label
11	handtas	vrouw	0.638115	Association
12	hoofddoek	vrouw	0.821153	Association
13	mascara	vrouw	0.766208	Association
14	pruik	vrouw	0.682367	Association
15	rouge	vrouw	0.823189	Association
16	jurk	vrouw	0.53669	Association
17	meisje	vrouw	0.43151	Label
18	haarstukje	vrouw	0.565204	Association
19	panty	vrouw	0.702756	Association
20	mantelpak	vrouw	0.656435	Association
1	vrouw	vrouw	0	Identity
2	vrouw	vrouw	0	Identity
3	vrouw	vrouw	0	Identity
4	vrouw	vrouw	0	Identity

Table 1.2. Continued.

Prime count (associations before division)	prime	target	Cosine similarity	Prime type (division of associative)
5	vrouw	vrouw	0	Identity
6	vrouw	vrouw	0	Identity
7	vrouw	vrouw	0	Identity
8	vrouw	vrouw	0	Identity
9	vrouw	vrouw	0	Identity
10	vrouw	vrouw	0	Identity
11	vrouw	vrouw	0	Identity
12	vrouw	vrouw	0	Identity
13	vrouw	vrouw	0	Identity
14	vrouw	vrouw	0	Identity
15	vrouw	vrouw	0	Identity
16	vrouw	vrouw	0	Identity
17	vrouw	vrouw	0	Identity
18	vrouw	vrouw	0	Identity
19	vrouw	vrouw	0	Identity
20	vrouw	vrouw	0	Identity
1	man	man	0	Identity
2	man	man	0	Identity
3	man	man	0	Identity
4	man	man	0	Identity
5	man	man	0	Identity
6	man	man	0	Identity
7	man	man	0	Identity
8	man	man	0	Identity
9	man	man	0	Identity
10	man	man	0	Identity
11	man	man	0	Identity
12	man	man	0	Identity
13	man	man	0	Identity
14	man	man	0	Identity
15	man	man	0	Identity
16	man	man	0	Identity
17	man	man	0	Identity
18	man	man	0	Identity
19	man	man	0	Identity
20	man	man	0	Identity

Table 2 List of models for the congruency analysis.

Model subset	v	A	T_{er}	DIC
0	C/T	C/T	C/T	-16537.9
1	C	C/T	C/T	-16549.3
2	C/T	C	C/T	-16485.2
3	C/T	C/T	C	-16080.9
4	C	C	C	-16547.9
5	T	C/T	C/T	-16515.5
6	C/T	T	C/T	-16601.7
7	C/T	C/T	T	-16110.6
8	T	T	T	-16147.8
9	C	C	C/T	-16496.3
10	C	C/T	C	-16114.7
11	C/T	C	C	-15890.1
12	T	T	C/T	-16588.0
13	T	C/T	T	-16074.3
14	C/T	T	T	-16213.7

Each of the models consists of three parameters (V, A, T_{er}) with the experimental factors (C = congruency, T = prime type) either fixed or left free per parameter. Letters present in the table (C, T, C/T) indicate the conditions over which the parameter was left free. Deviance information criterion scores (DIC) indicates the goodness of model fit. The model with the lower DIC is considered to be the model winner.

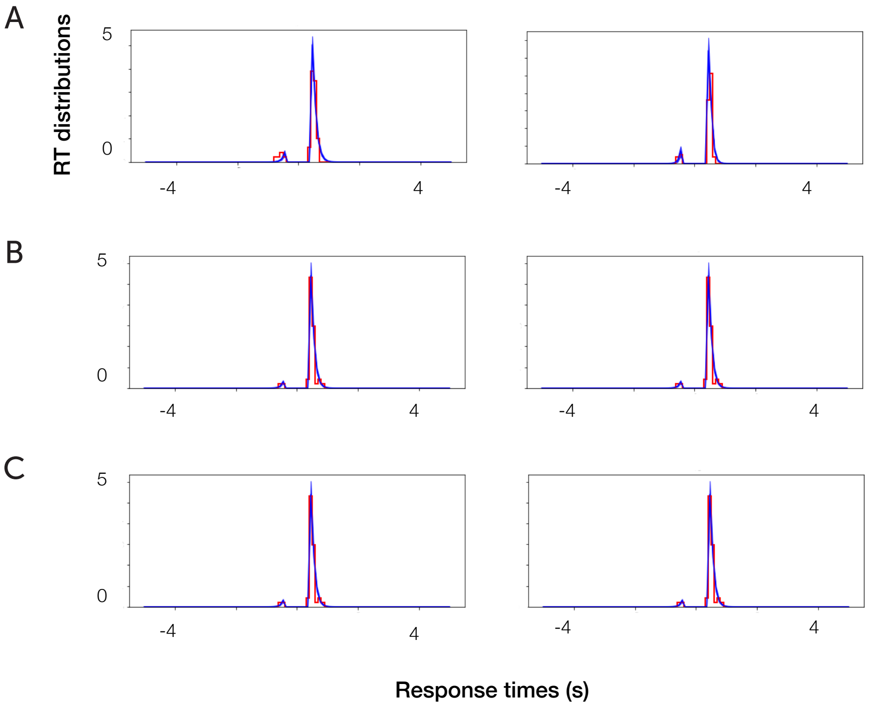
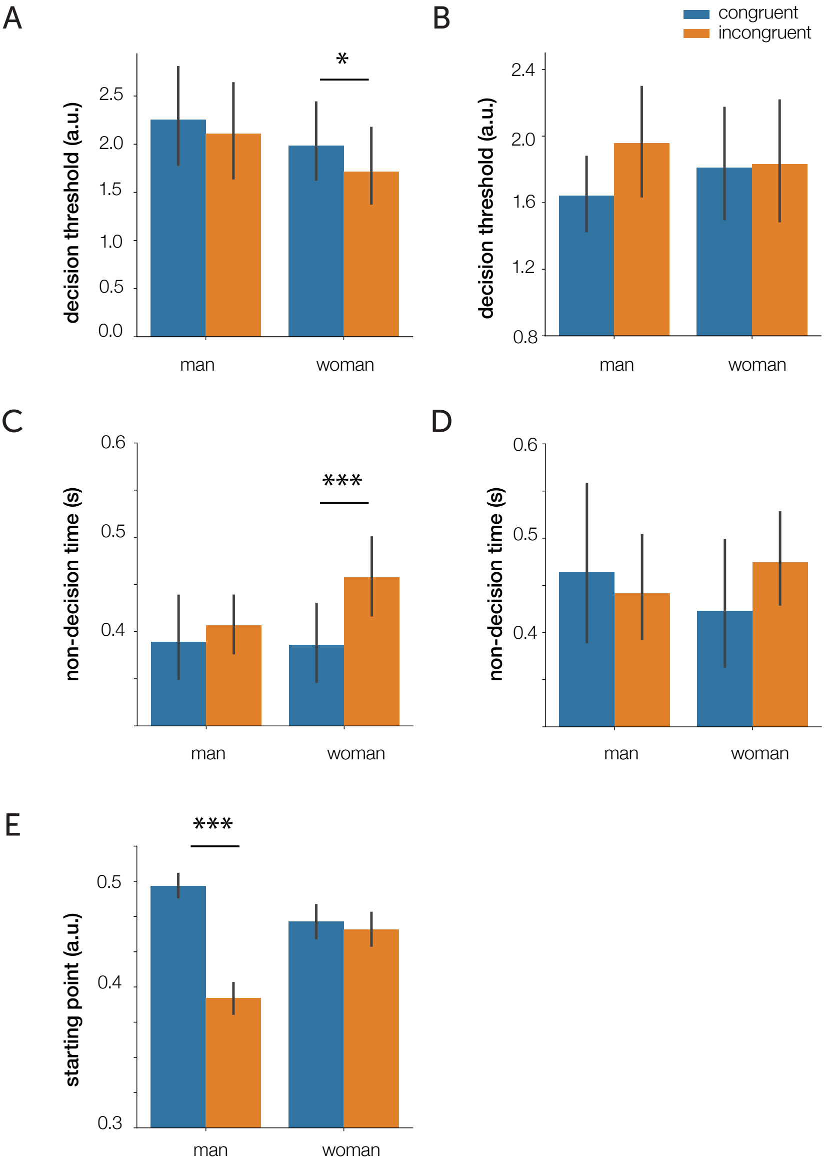


Figure 1: Congruency analysis: Observed RT distributions (red lines) and predicted posteriors (blue line) for the drift rate for a representative subject (subject 10(A left – congruent associative, A right – congruent identity; B left – incongruent associative, B right – incongruent identity; C left – neutral associative, C right – neutral identity).

SUPPLEMENT FOR CHAPTER 3



Supplementary Figure 1: HDDM parameter estimates: decision threshold for classification **(A)** and verification **(B)** tasks; non-decision time for classification **(C)** and verification **(D)** tasks and starting point for verification task **(E).** Error bars are standard error of the mean. * $p < 0.05$, *** $p < 0.001$

Supplementary Information: Drift Diffusion Model Parameters

Decision threshold (A)

We found an interaction between task and congruency as well as an interaction between task and target (see Supp. Table 1). For the classification task (Supp. Figure1, A), we did not find any differences between congruent and incongruent pairs with male targets: *t* (21) = 1.64, *p* = .058). We found increased threshold for congruent in comparison to incongruent pairs with female targets *t* (21) = 2.46, *p* = .011).
For the verification task (Supp. Figure 1, B) we did not perform any follow-up tests since we did not find any significant effect (neither main effects of congruency or target nor interaction between congruency and target).

Non-decisional component (Ter)

We found an interaction between task and target (see Supp. Table 1). For the classification task (Supp. Figure1, C), incongruent trials in comparison to congruent ones with female faces resulted in slower RT: *t* (21) = 4.00, *p* < .001. We did not find significant difference between congruent vs. incongruent pairs with male faces: *t* (21) = 1.02, *p* = .15).
For the verification task (Supp. Figure 1, D), we did not perform any follow-up tests since we did not find any significant effect (neither main effects of congruency or target nor interaction between congruency and target).

Starting point (z)

We did not get any effects of interest on the starting point (see *Model Convergence and Model Fit*) for the classification task, but only for the verification task.
For the verification task (Supp. Figure 1, E), there was an increased starting point for congruent in comparison to incongruent pairs with male targets: *t* (21) = 10.68, *p* < .001. We did not find significant differences between congruent and incongruent pairs with female targets: *t* (21) = .71, *p* = .24.

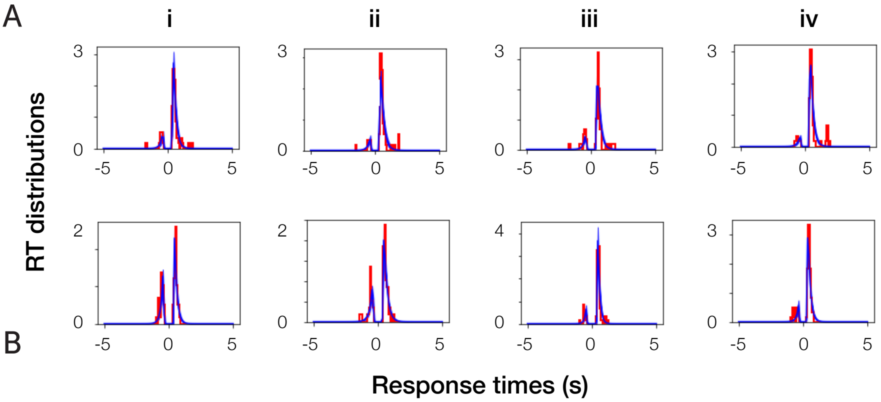
In sum, HDDM analysis showed that two different sets of HDDM parameters underlie the RTs and accuracies in the classification and verification tasks. While in the classification task drift rate, decision threshold and non-decisional component are at play, in the verification task, only the drift rate and starting point seem to play a role.

Supplementary Table 1 Repeated-measures ANOVA (2-way) results for decision threshold, non-decisional component and starting.

Decision Threshold	*F(1,21)*	*p*	η^2
Task	*2.462*	*.132*	*.043*
Congruency	*.123*	*.729*	4.072×10^{-4}
Target	*5.356*	*.031**	*.025*
Task * Congruency	*8.128*	*.010**	*.036*
Task * Target	*6.791*	*.017**	*.032*
Congruency * Target	*3.131*	*.091*	*.011*
Task * Congruency * Target	*.765*	*.392*	*.002*
Non-decisional component			
Task	*4.341*	*.050*	*.054*
Congruency	*3.875*	*.062*	*.028*
Target	*.409*	*.529*	*.003*
Task * Congruency	*1.792*	*.195*	*.007*
Task * Target	*4.392*	*.048**	*.006*
Congruency * Target	*8.844*	*.007***	*.034*
Task * Congruency * Target	*.191*	*.666*	8.048×10^{-4}
Starting point			
Congruency	*59.818*	*<.001****	*.367*
Target	*7.638*	*.012**	*.028*
Congruency * Target	*47.185*	*< .001****	*.276*

Factors tested were congruency (levels: congruent, incongruent) and target (levels: male, female).
* $p < .05$; ** $p < .01$; *** $p < .001$. CL – classification task, VR – verification task.

HDDM Model fit



Supplementary Figure 2: Observed RT distributions (red lines) and predicted posteriors (blue line) for the drift rate in classification **(A)** and verification **(B)** tasks (i – prime: man, target: man; ii – prime: man, target: woman; iii – prime: woman, target: man; iv – prime: woman, target: woman).

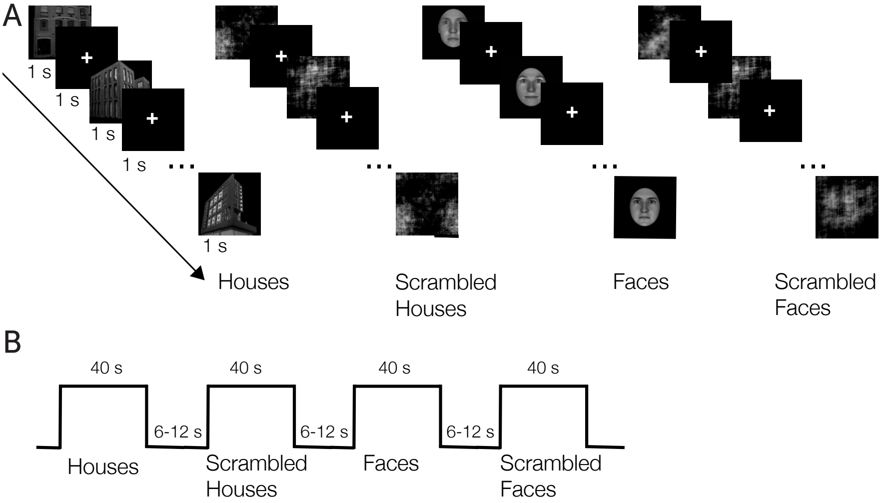
Supplementary Table 2 Total model list for analysis of priming effects on face gender.

Model set	Task	V	A	Ter	z	DIC
0	1	PT	PT	PT	-	-210.26
0	2	PT	PT	PT	-	753.77
1	1	P	PT	PT	-	-38.97
1	2	P	PT	PT	-	891.01
2	1	T	PT	PT	-	-120.21
2	2	T	PT	PT	-	905.70
3	1	PT	P	PT	-	-168.84
3	2	PT	P	PT	-	771.10
4	1	PT	T	PT	-	-177.75
4	2	PT	T	PT	-	756.38
5	1	PT	PT	P	-	-52.67
5	2	PT	PT	P	-	940.76
6	1	PT	PT	T	-	-67.89
6	2	PT	PT	T	-	961.69
7	1	P	P	PT	-	-80.29

Supplementary Table 2 Continued.

Model set	Task	V	A	Ter	z	DIC
7	2	P	P	PT	-	883.88
8	1	P	T	PT	-	-59.82
8	2	P	T	PT	-	915.73
9	1	T	P	PT	-	-128.29
9	2	T	P	PT	-	915.73
10	1	T	T	PT	-	-156.39
10	2	T	T	PT	-	890.72
11	1	P	P	P	-	153.56
11	2	P	P	P	-	1293.99
12	1	P	T	P	-	148.89
12	2	P	T	P	-	1309.94
13	1	T	P	P	-	84.21
13	2	T	P	P	-	1308.24
14	1	T	T	P	-	66.96
14	2	T	T	P	-	1295.77
15	1	P	P	T	-	133.25
15	2	P	P	T	-	1345.79
16	1	P	T	T	-	165.96
16	2	P	T	T	-	1321.99
17	1	T	P	T	-	119.74
17	2	T	P	T	-	1336.49
18	1	T	T	T	-	96.90
18	2	T	T	T	-	1420.02
19	1	PT	PT	PT	P	-199.92
19	2	PT	PT	PT	P	740.89
20	1	PT	PT	PT	T	-208.32
20	2	PT	PT	PT	T	757.09
21	1	PT	PT	PT	PT	-210.14
21	2	PT	PT	PT	PT	710.31

Each model is comprised of three parameters (V, A, T_{er}) with the experimental factors (P – prime, T – target stimulus) either fixed or left free per parameter. Letters present in the table (P, T, P/T) indicate the conditions over which the parameter was left free. Deviance information criterion scores (DIC) indicates the goodness of model fit. The model with the lowest DIC is considered to be the model winner.



Supplementary Figure 3: A. Task structure for the Localizer. Each of four blocks - faces, houses, scrambled faces and scrambled houses - started with the fixation cross presented for 1 s followed by 1 s of stimulus presentation. **B**. The overall timing of the localizer. Each of these blocks had a total duration of 40 s with jittered 6-12 s of break in between.

Supplementary Table 3. BOLD contrasts for main effects (congruency, target, interaction between target and congruency) for the whole brain analysis, CL – classification, VR – verification. ns, not-significant.

Task	Contrast	*Whole brain*
CL	Congruency	*ns*
VR	Congruency	*ns*
CL	Target	*ns*
VR	Target	*ns*
CL	Target*Congruency	*ns*
VR	Target*Congruency	*ns*

Supplementary Table 4. BOLD contrasts for exploratory effects for the whole brain analysis in the verification task. ns, not-significant.

Contrast	*Whole brain*
Congruent female > incongruent female	*ns*
Incongruent female > congruent female	*ns*
Congruent male > incongruent male	*ns*
Incongruent male > congruent male	*ns*

Supplementary Table 5. Repeated-measures ANOVA using BOLD activation within the FFA for main effects (congruency, target, interaction between target and congruency)

Within Subjects Effects						
Cases	Sum of Squares	df	Mean Square	F	p	η^2
task	3.138	1	3.138	0.049	0.827	0.002
Residuals	1341.411	21	63.877			
cong	1.199	1	1.199	0.551	0.466	6.890×10^{-4}
Residuals	45.691	21	2.176			
target	3.419	1	3.419	1.915	0.181	0.002
Residuals	37.484	21	1.785			
task ✻ cong	13.766	1	13.766	3.425	0.078	0.008
Residuals	84.404	21	4.019			
task ✻ target	2.116	1	2.116	1.009	0.327	0.001
Residuals	44.036	21	2.097			
cong ✻ target	0.920	1	0.920	0.203	0.657	5.288×10^{-4}
Residuals	94.967	21	4.522			
task ✻ cong ✻ target	4.917	1	4.917	1.643	0.214	0.003
Residuals	62.862	21	2.993			

Note. Type III Sum of Squares

Supp. Table 6. Correlation results between drift rate and BOLD for congruent effect and gender-specific congruency effects.

Task: parameter: contrast	**Contrast (BOLD)**	***Whole brain***
CL: drift rate: Congruent > incongruent	Congruent > incongruent	*ns*
VR: drift rate: Congruent female > incongruent female	Congruent female > incongruent female	*ns*
VR: drift rate: Incongruent male > congruent male	Incongruent male > congruent male	***.04***

CL – classification, VR – verification. ns, not significant.

SUPPLEMENT FOR CHAPTER 4

Supp. Table 1. List of associatively related to face and house concept and neutral words.

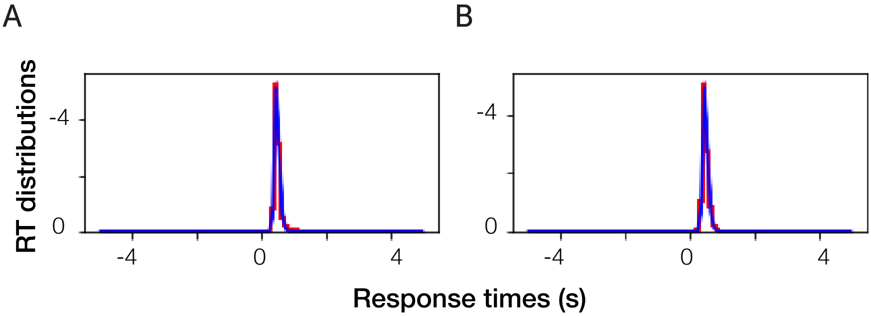
Words related to face concept	Words related to house concept	Neutral words
Stoppelbaard (stubble)	Plafond (ceiling)	Zalm (salm)
Schedel (skull)	Iglo (iglo)	Knopen (knot)
Pukkel (pimple)	Toren (tower)	Knoflook (garlic)
Tandvlees (gums)	Zaal (room)	Rundvlees (beef)
Puist (pimple)	Flatgebouw (appartment building)	Kaneel (cinnamon)
Baard (beard)	Boomhut (treehouse)	Pindakaas (peanut butter)
Gebit (teeth)	Blokhut (cabin)	Gember (ginger)
Snor (mustache)	Hemelbed (canopy bed)	Geitenkaass (goat cheese)
Flaporen (floppy ears)	Koelkast (refrigerator)	Kabeljauw (codfish)
Bovenlip (upper lip)	Paleis (palace)	Rekenmachine (calculator)
Onderkin (double chin)	Oprit (ramp)	Paprika (paprika)
Spoeten (freckles)	Eettafel (dining table)	Algen (algae)
Kin (chin)	Schouw (fireplace)	Vruchtensap (fruit juice)
Oogbal (eyeball)	Duivenhok (loft)	Drum (drum)
Hals (neck)	Afdak (shed)	Rozijnen (raisins)
Voorhoofd (forehead)	Openhaard (fireplace)	Vanilla (vanille)
Ooglid (eyelid)	Schoorsteen (chimney)	Haardroger (hairdryer)
Wang (cheek)	Wastobbe (washtub)	Margriet (flower)
Neusvleugel (nostril)	Keukendeur (kitchen door)	Steenkool (coal)
Oorlei (earlobe)	Voordeur (front door)	Ansjovis (anchovy)

Supp. Table 2. Total model list for the HDDM analysis.

Model set	V	A	Ter	DIC
0	fixed	fixed	**free**	-16421.4
1	fixed	**free**	fixed	-16462.9
2	**free**	fixed	fixed	-16450.7
3	fixed	**free**	**free**	-16453.7
4	**free**	fixed	**free**	-16442.2
5	**free**	**free**	fixed	-16440
6	**free**	**free**	**free**	-16461.7

Each of the models consists of three parameters (V, A, T_{er}) with the experimental factor (congruency) either fixed or left free per parameter. Deviance information criterion scores (DIC) indicates the goodness of model fit. The model with the lower DIC is considered to be the model winner.

HDDM Model fit



Supp. Figure 1: Observed RT distributions (red lines) and predicted posteriors (blue line) for the drift rate in congruent and incongruent condition for one representative subject.

Supp. Table 3. Spearman correlations for ERF and HDDM parameters (A: decision threshold; V: drift rate; Ter: non-decision component) for congruent (C) > incongruent (IC) conditions.

		C>IC ERF	C>IC A	C>IC V	C>IC Ter
C>IC ERF	Spearman’s rho	—			
	p-value	—			
C>IC A	Spearman’s rho	0.266	—		
	p-value	0.219	—		
C>IC V	Spearman’s rho	0.093	0.776***	—	
	p-value	0.672	< .001	—	
C>IC Ter	Spearman’s rho	-0.172	-0.742***	-0.366	—
	p-value	0.431	< .001	0.087	—

* p < .05, ** p < .01, *** p < .001

SUPPLEMENT FOR CHAPTER 5

We performed an additional analysis for the patients in temporal and frontal groups excluding one patient from each group due to the small lesion volume as it was suggested by an anonymous reviewer. The results are presented in the Supp. Table 1 and Supp. Figure 1. For this analysis we use the same methods as described in the main body of the paper. The results of the analysis with excluded subjects do not differ from those reported in the main analysis, particularly, the effects of interest, unrelated > related (in both drift rate and decision threshold). Patients in both groups had increased drift rate and threshold for unrelated in comparison to related conditions. The strength of the evidence for each Bayesian test did not drop down below “positive”. We however did not find the effects of participant type on the non-decision component, which nevertheless did not affect our inferences about contributions of lexico-semantic vs. control systems.

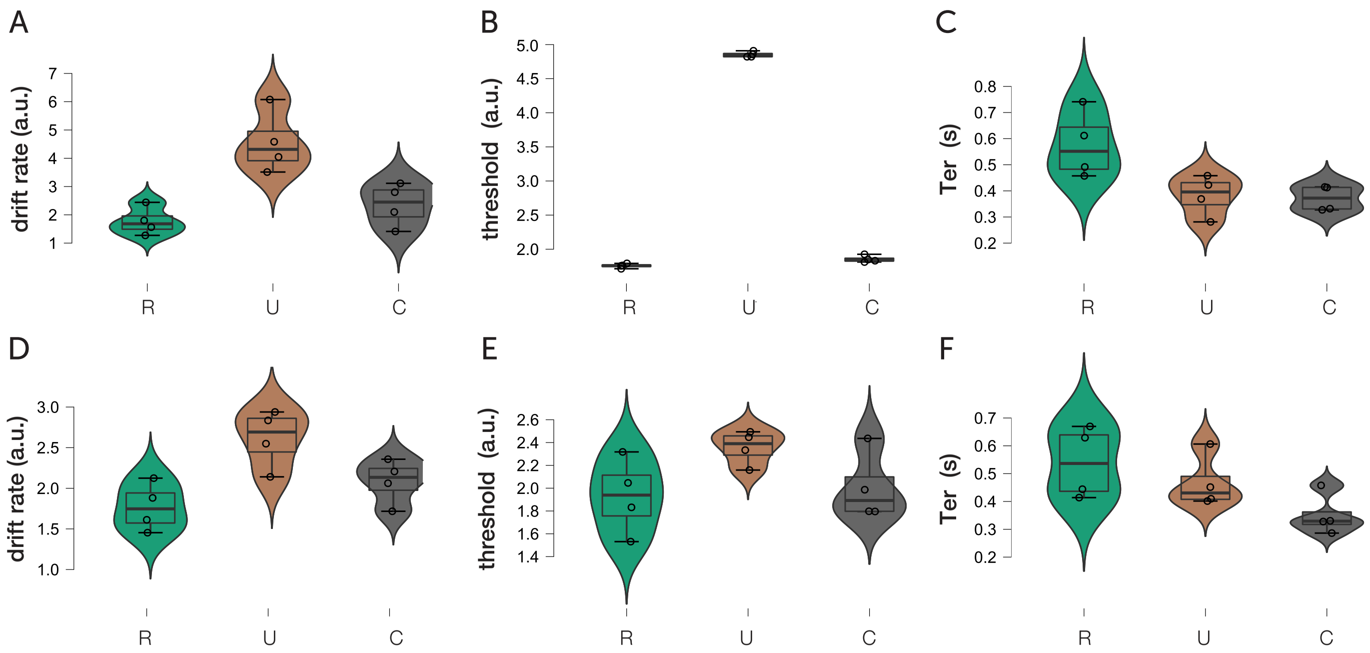
Supp. Table 1. Deviance information criterion (DIC) for each model for patients in frontal and temporal groups (4 patients per group with patient 4 and patient 9 excluded).

subject type/model	1	2	3	4	5	6	7
Frontal patients	351.69	364.83	371.48	389.80	530.20	443.97	396.48
Temporal patients	747.72	772.72	758.17	753.44	884.23	792.26	763.33

For both patients in temporal and frontal groups, the best winning model had drift rate, decision threshold and non-decision time free (Model 1, Supp. Figure 1).

Drift rate

Supp. Figure 1, A, D. shows the drift rate estimates for each participant as a function of subject type (frontal, temporal) and relatedness (related, unrelated, congruent). A JZS Bayes factor ANOVA with default prior scales revealed that the best-fitting model consisted of main effects of word-picture relatedness, type of subject and an interaction between word-picture relatedness and type of subject. This model was preferred over the null model, BF_{10} = 17070.985 and over the second-best model that included main effects of word-picture relatedness and subject type, BF_{10} = 612.125.



Supp. Figure 1. Model parameter estimates (drift rate **(A)**, decision threshold **(B)** and non-decision component i.e. Ter **(C)** for patients in frontal group (upper row) and temporal group (lower row).

Further, we performed Bayesian paired samples t-tests for subjects with frontal and temporal lesions separately. For patients with temporal lesions, there was “positive” evidence in support of increased drift rate in unrelated relative to related conditions (unrelated > related: BF_{10} = 6.62). For the patients with frontal lesions, we found “strong” evidence in favor of increased drift rate in unrelated relative to related conditions (unrelated > related: BF_{10} = 33.03).

For completeness, we report the result for the additional contrasts. For patients with temporal lesions, we found “anecdotal” evidence in favor of a difference between related vs. congruent pairs (related vs. congruent: BF_{10} = 2.13) and “positive” evidence in favor of increased rate in unrelated in comparison to congruent conditions, (unrelated > congruent: BF_{10} = 8.35). For the patients with frontal lesions, we found “strong” evidence in favor of increased drift rate in unrelated in comparison to congruent conditions (unrelated > congruent: BF_{10} = 13.49), whereas related and congruent conditions showed “anecdotal” evidence supporting the condition differences (related vs. congruent: BF_{10} = 1.43).

In sum, we found that both temporal and frontal patients had increased drift rate for unrelated in comparison to related word-picture pairs.

Decision threshold:

A JZS Bayes factor ANOVA with default prior scales revealed that the best-fitting model consisted of main effects of word-picture relatedness, type of subject and an interaction between word-picture relatedness and type of subject. This model was preferred over the null model, BF_{10} = 6.346e+9 and over the second-best model that included main effects of word-picture relatedness and subject type, BF_{10} = 231.578.

For the patients with frontal lesions, there was “very strong” evidence in favor of higher decision threshold in unrelated vs. related pairs (unrelated > related pairs: BF_{10} = 4134). Thus, the direction of the effect was reversed in comparison to the controls. For patients with temporal lesions, we found “positive” evidence in favor of increased decision threshold in unrelated compared to related conditions (unrelated > related: BF_{10} = 3.17).

For patients with frontal lesions, there was “very strong” evidence in favor of increased decision threshold for unrelated in comparison to congruent conditions (unrelated > congruent: BF_{10} = 121691). By contrast, the evidence supporting the difference between related and congruent conditions was “anecdotal”, (related vs. congruent: BF_{10} = 2.45). We observed a similar pattern for patients with temporal lesions. We found "anecdotal" evidence in favor of the difference between related vs. congruent pairs (related vs. congruent: BF_{10} = 0.43), and "positive" evidence in favor of increased decision threshold in unrelated vs. congruent conditions (unrelated > congruent: BF_{10} = 3.96).

The decision threshold was lower for related relative to unrelated word-picture pairs in both patient groups when compared to controls.

Non decision component

A JZS Bayes factor ANOVA with default prior scales revealed that the model-winner included main effect of relatedness. This model was preferred over the null model BF_{10} = 86.67 and over the second-best model that included main effects of word-picture relatedness and subject type BF_{10} = 33.01.

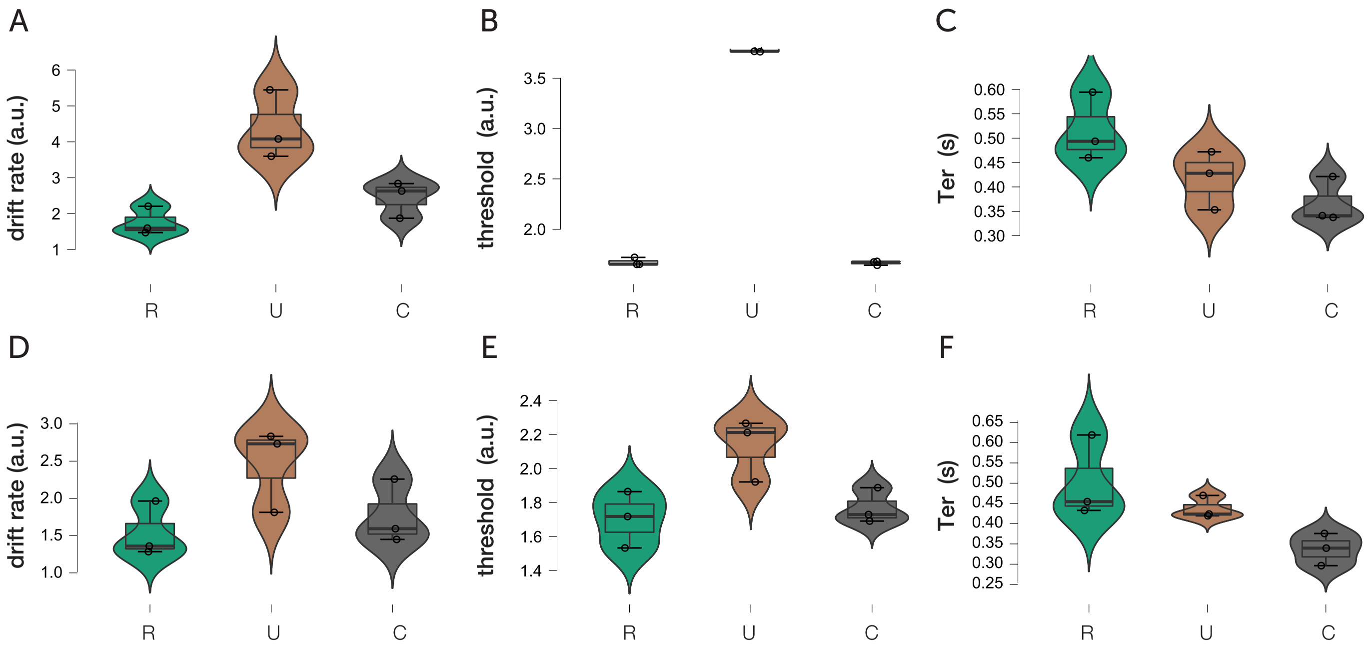
Since the best model did not indicate the effect of patient type on task performance, we did not test the effects of semantic relatedness further.

We performed an additional analysis for the patients in temporal and frontal groups excluding patients from each group due to mixed parietal contributions. This resulted in three patients in each group: temporal group (patients 1, 2, 3) and frontal group (patients 6, 7, 10). The results are presented in Supp. Table 2 and Supp. Figure 2. For this analysis we use the same methods as described in the main body of the paper. For patients in the frontal group, both drift rate and decision threshold were increased for unrelated in comparison to related condition. Strength of evidence for Bayesian test was no lower than "positive". For patients in the temporal group, we found only "anecdotal" evidence in favor of increased drift rate and decision threshold for unrelated in comparison to related conditions.

Supp. Table 2. Deviance information criterion (DIC) for each model for patients in frontal and temporal groups.

subject type/model	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Frontal patients	-129.70	-97.00	-102.99	-78.06	41.56	-34.45	-77.12
Temporal patients	417.45	464.55	423.60	431.49	551.33	502.66	435.52

For both patients in temporal and frontal groups, the best winning model had drift rate, decision threshold and non-decision time free (Model 1, Supp. Figure 2).



Supp. Figure 2: Model parameter estimates (drift rate **(A)**, decision threshold **(B)** and non-decision component i.e. Ter **(C)** for patients in frontal group (upper row) and temporal group (lower row).

Drift rate

Supp. Figure 2, A, D. shows the drift rate estimates for each participant as a function of subject type (frontal, temporal) and relatedness (related, unrelated, congruent). A JZS Bayes factor ANOVA with default prior scales revealed that the best-fitting model consisted of main effects of word-picture relatedness, type of subject and an interaction between word-picture relatedness and type of subject. This model was preferred over the null model, BF_{10} = 349.164 and over the second-best model that included main effects of word-picture relatedness and subject type, BF_{10} = 126.22.

Further, we performed Bayesian paired samples t-tests for subjects with frontal and temporal lesions separately. For patients with temporal lesions, there was "anecdotal" evidence in support of increased drift rate in unrelated relative to related conditions (unrelated > related: BF_{10} = 2.98). For the patients with frontal lesions, we found "positive" evidence in favor of increased drift rate in unrelated relative to related conditions (unrelated > related: BF_{10} = 9.99).

For completeness, we report the result for the additional contrasts. For patients with temporal lesions, we found "anecdotal" evidence in favor of a difference between related vs. congruent pairs (related vs. congruent: BF_{10} = 1.88) and "anecdotal" evidence in favor of increased rate in unrelated in comparison to congruent conditions, (unrelated > congruent: BF_{10} = 2.59). For the patients with frontal lesions, we found "positive" evidence in favor of increased drift rate in unrelated in comparison to congruent conditions (unrelated > congruent: BF_{10} = 4.66), whereas related and congruent conditions showed "anecdotal" evidence supporting the condition differences (related vs. congruent: BF_{10} = 1.40).

In sum, we found that both temporal and frontal patients had increased drift rate for unrelated in comparison to related word-picture pairs.

Decision threshold:

A JZS Bayes factor ANOVA with default prior scales revealed that the best-fitting model consisted of main effects of word-picture relatedness, type of subject and an interaction between word-picture relatedness and type of subject. This model was preferred over the null model, BF_{10} = 2.55 x 10^6 and over the second-best model that included main effects of word-picture relatedness and subject type, BF_{10} = 56.49

For the patients with frontal lesions, there was "strong" evidence in favor of higher decision threshold in unrelated vs. related pairs (unrelated > related pairs: BF_{10} =

86.90). Thus, the direction of the effect was reversed in comparison to the controls. For patients with temporal lesions, we found "anecdotal" evidence in favor of increased decision threshold in unrelated compared to related conditions (unrelated > related: $BF_{10} = 2.82$).

For patients with frontal lesions, there was "very strong" evidence in favor of increased decision threshold for unrelated in comparison to congruent conditions (unrelated > congruent: $BF_{10} = 255.39$). By contrast, the evidence supporting the difference between related and congruent conditions was "anecdotal", (related vs. congruent: $BF_{10} = 0.50$). We observed a similar pattern for patients with temporal lesions. We found "anecdotal" evidence in favor of the difference between related vs. congruent pairs (related vs. congruent: $BF_{10} = 0.50$), and "positive" evidence in favor of increased decision threshold in unrelated vs. congruent conditions (unrelated > congruent: $BF_{10} = 5.75$).

The decision threshold was lower for related relative to unrelated word-picture pairs in both patient groups when compared to controls.

Non decision component

A JZS Bayes factor ANOVA with default prior scales revealed that the model-winner included main effect of relatedness. This model was preferred over the null model $BF_{10} = 46.77$ and over the second-best model that included main effects of word-picture relatedness and subject type $BF_{10} = 19.22$.

Since the best model did not indicate the effect of patient type on task performance, we did not test the effects of semantic relatedness further.

To investigate whether lesions in STG contributed to the parameter estimates, we performed a Bayesian correlation between percentage of damage in STG and parameter estimates in the unrelated condition (drift rate and decision threshold). We performed Bayesian correlations for both supplementary analysis 1 (four patients in both temporal and frontal groups) and supplementary analysis 2 (three patients in both temporal and frontal groups). We report the results of Bayesian correlations in Supplementary Table 3 (for supplementary analysis 1) and Supplementary Table 4 (for supplementary analysis 2). In short, we did not find any positive evidence ($BF_{10} > 3$) in favor of a correlation between HDDM parameters (drift rate, decision threshold) in the unrelated condition and percentage of damage in STG.

Bibliography

Nederlandse samenvatting

Acknowledgments

Curriculum Vitae

Publications

Research Data Management

Donders Graduate School for Cognitive Neuroscience

Bibliography

Amado, C., Kovács, P., Mayer, R., Ambrus, G. G., Trapp, S., & Kovács, G. (2018). Neuroimaging results suggest the role of prediction in cross-domain priming. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-28696-0

Anders, R., Ries, S., van Maanen, L., & Alario, F.-X. (2015). Evidence accumulation as a model for lexical selection. *Cognitive Psychology*, *82*, 57–73. https://doi.org/10.1016/j.cogpsych.2015.07.002

Anders, R., Riès, S., Van Maanen, L., & Alario, F.-X. (2017). Lesions to the left lateral prefrontal cortex impair decision threshold adjustment for lexical selection. *Cognitive Neuropsychology*, *34*(1–2), 1–20. https://doi.org/10.1080/02643294.2017.1282447

Anderson, D. E., Serences, J. T., Vogel, E. K., & Awh, E. (2014). Induced alpha rhythms track the content and quality of visual working memory representations with high temporal precision. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *34*(22), 7587–7599. https://doi.org/10.1523/JNEUROSCI.0293-14.2014

Armel, K. C., Beaumel, A., & Rangel, A. (2008). *Biasing simple choices by manipulating relative visual attention. Judgment and Decision Making* (Vol. 3).

Badre, D., Poldrack, R. A., Paré-Blagoev, E. J., Insler, R. Z., & Wagner, A. D. (2005). Dissociable Controlled Retrieval and Generalized Selection Mechanisms in Ventrolateral Prefrontal Cortex. *Neuron*, *47*(6), 907–918. https://doi.org/10.1016/j.neuron.2005.07.023

Baldo, J. V., Arévalo, A., Patterson, J. P., & Dronkers, N. F. (2013). Grey and white matter correlates of picture naming: Evidence from a voxel-based lesion analysis of the Boston Naming Test. *Cortex*, *49*(3), 658–667. https://doi.org/10.1016/J.CORTEX.2012.03.001

Bannert, M. M., & Bartels, A. (2013). Decoding the yellow of a gray banana. *Current Biology*, *23*(22), 2268–2272. https://doi.org/10.1016/j.cub.2013.09.016

Barsalou, L. W. (2008). Grounded Cognition. *Annual Review of Psychology*, *59*(1), 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Bitzer, S., Park, H., Blankenburg, F., & Kiebel, S. J. (2014). Perceptual decision making: drift-diffusion model is equivalent to a Bayesian model. *Frontiers in Human Neuroscience*, *8*, 102. https://doi.org/10.3389/fnhum.2014.00102

Blanz, V., & Vetter, T. (1999). A morphable model for the synthesis of 3D faces. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques - SIGGRAPH '99* (pp. 187–194). New York, New York, USA: ACM Press. https://doi.org/10.1145/311535.311556

Bogacz, R., Hu, P. T., Holmes, P. J., & Cohen, J. D. (2010). Do humans produce the speed-accuracy trade-off that maximizes reward rate? *Quarterly Journal of Experimental Psychology (2006)*, *63*(5), 863–891. https://doi.org/10.1080/17470210903091643

Bogacz, R., Wagenmakers, E. J., Forstmann, B. U., & Nieuwenhuis, S. (2010). The neural basis of the speed-accuracy tradeoff. *Trends in Neurosciences*. Elsevier Current Trends. https://doi.org/10.1016/j.tins.2009.09.002

Boroditsky, L. (2001). Does Language Shape Thought?: Mandarin and English Speakers' Conceptions of Time. *Cognitive Psychology*, *43*(1), 1–22. https://doi.org/10.1006/cogp.2001.0748

Bourguignon, N. J. (2014). A rostro-caudal axis for language in the frontal lobe: The role of executive control in speech production. *Neuroscience & Biobehavioral Reviews*, *47*, 431–444. https://doi.org/10.1016/J.NEUBIOREV.2014.09.008

Boutonnet, B., & Lupyan, G. (2015). Words Jump-Start Vision: A Label Advantage in Object Recognition. *Journal of Neuroscience*, *35*(25), 9329–9335. https://doi.org/10.1523/JNEUROSCI.5111-14.2015

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a New Set of 480 Normative Photos of Objects to Be Used as Visual Stimuli in Cognitive Research. *PLoS ONE*, *5*(5), e10773. https://doi.org/10.1371/journal.pone.0010773

Brouwer, G. J., & Heeger, D. J. (2013). Categorical clustering of the neural representation of color. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *33*(39), 15454–15465. https://doi.org/10.1523/JNEUROSCI.2472-13.2013

Brunton, B. W., Botvinick, M. M., & Brody, C. D. (2013). Rats and humans can optimally accumulate evidence for decision-making. *Science (New York, N.Y.)*, *340*(6128), 95–98. https://doi.org/10.1126/science.1233912

Burnham, K. P., Anderson, D. R., & Burnham, K. P. (2002). *Model selection and multimodel inference : a practical information-theoretic approach*. Springer.

Campanella, F., Crescentini, C., Mussoni, A., & Skrap, M. (2013). Refractory semantic access dysphasia resulting from resection of a left frontal glioma. *Neurocase*, *19*(1), 27–35. https://doi.org/10.1080/13554794.2011.654212

Campanella, F., & Shallice, T. (2011). Refractoriness and the healthy brain: A behavioural study on semantic access. *Cognition*, *118*(3), 417–431. https://doi.org/10.1016/j.cognition.2010.08.005

Carroll, N. C., & Young, A. W. (2005). Priming of emotion recognition. *The Quarterly Journal of Experimental Psychology Section A*, *58*(7), 1173–1197. https://doi.org/10.1080/02724980443000539

Cavanagh, J. F., Wiecki, T. V, Cohen, M. X., Figueroa, C. M., Samanta, J., Sherman, S. J., & Frank, M. J. (2011). Subthalamic nucleus stimulation reverses mediofrontal influence over decision threshold. *Nature Neuroscience*, *14*(11), 1462–1467. https://doi.org/10.1038/nn.2925

Chao, L. L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neuroscience*, *2*(10), 913–919. https://doi.org/10.1038/13217

Chen, Q., Middleton, E., & Mirman, D. (2018). Words fail: Lesion-symptom mapping of errors of omission in post-stroke aphasia. *Journal of Neuropsychology*. https://doi.org/10.1111/jnp.12148

Chiou, R., Humphreys, G. F., Jung, J., & Lambon Ralph, M. A. (2018). Controlled semantic cognition relies upon dynamic and flexible interactions between the executive 'semantic control' and hub-and-spoke 'semantic representation' systems. *Cortex*, *103*, 100–116. https://doi.org/10.1016/J.CORTEX.2018.02.018

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: a parallel distributed processing account of the Stroop effect. *Psychological Review*, *97*(3), 332–361. https://doi.org/10.1037/0033-295x.97.3.332

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428. https://doi.org/10.1037/0033-295X.82.6.407

Contreras, J. M., Banaji, M. R., & Mitchell, J. P. (2013). Multivoxel Patterns in Fusiform Face Area Differentiate Faces by Sex and Race. *PLoS ONE*, *8*(7), e69684. https://doi.org/10.1371/journal.pone.0069684

Copland, D. A., De Zubicaray, G. I., McMahon, K., Wilson, S. J., Eastburn, M., & Chenery, H. J. (2003). Brain activity during automatic semantic priming revealed by event-related functional magnetic resonance imaging. *NeuroImage*, *20*(1), 302–310. https://doi.org/10.1016/S1053-8119(03)00279-9

Coutrot, A., Binetti, N., Harrison, C., Mareschal, I., & Johnston, A. (2016). Face exploration dynamics differentiate men and women. *Journal of Vision*, *16*(14), 16. https://doi.org/10.1167/16.14.16

Crutch, S. J., & Warrington, E. K. (2005). Gradients of semantic relatedness and their contrasting explanations in refractory access and storage semantic impairments. *Cognitive Neuropsychology*, *22*(7), 851–876. https://doi.org/10.1080/02643290442000374

Davey, J., Thompson, H. E., Hallam, G., Karapanagiotidis, T., Murphy, C., De Caso, I., … Jefferies, E. (2016). Exploring the role of the posterior middle temporal gyrus in semantic cognition: Integration of anterior temporal lobe with executive processes. *NeuroImage*, *137*, 165–177. https://doi.org/10.1016/j.neuroimage.2016.05.051

De Deyne, S., Navarro, D. J., & Storms, G. (2013). Better explanations of lexical and semantic cognition using networks derived from continued rather than single-word associations. *Behavior Research Methods*, *45*(2), 480–498. https://doi.org/10.3758/s13428-012-0260-7

De Houwer, J., Hermans, D., Rothermund, K., & Wentura, D. (2002). Affective priming of semantic categorisation responses. *COGNITION AND EMOTION*, *16*(5), 643–666. https://doi.org/10.1080/02699930143000419

de Lange, F. P., Heilbron, M., & Kok, P. (2018). How Do Expectations Shape Perception? *Trends in Cognitive Sciences*. Elsevier Ltd. https://doi.org/10.1016/j.tics.2018.06.002

de Zubicaray, G., Arciuli, J., & McMahon, K. (2013). Putting an "end" to the motor cortex representations of action words. *Journal of Cognitive Neuroscience*, *25*(11), 1957–1974. https://doi.org/10.1162/jocn_a_00437 de Zubicaray, Hansen, S., & McMahon, K. L. (2013). Differential processing of thematic and categorical conceptual relations in spoken word production. *Journal of Experimental Psychology: General*, *142*(1), 131–142. https://doi.org/10.1037/a0028717

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, *41*(6), 391–407. https://doi.org/10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, *104*(4), 801–838. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9337631

Dell, Martin, N., & Schwartz, M. F. (2007). A Case-Series Test of the Interactive Two-step Model of Lexical Access: Predicting Word Repetition from Picture Naming. *Journal of Memory and Language*, *56*(4), 490. https://doi.org/10.1016/J.JML.2006.05.007

Demb, J. B., Desmond, J. E., Wagner, A. D., Vaidya, C. J., Glover, G. H., & Gabrieli, J. D. (1995). Semantic encoding and retrieval in the left inferior prefrontal cortex: a functional MRI study of task difficulty and process specificity. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *15*(9), 5870–5878. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7666172

Domenech, P., & Dreher, J.-C. (2010). Decision Threshold Modulation in the Human Brain. *Journal of Neuroscience*, *30*(43), 14305–14317. https://doi.org/10.1523/JNEUROSCI.2371-10.2010

Donkin, C., Brown, S., Heathcote, A., & Wagenmakers, E.-J. (2011). Diffusion versus linear ballistic accumulation: different models but the same conclusions about psychological processes? *Psychonomic Bulletin & Review*, *18*(1), 61–69. https://doi.org/10.3758/s13423-010-0022-4

Dricot, L., Sorger, B., Schiltz, C., Goebel, R., & Rossion, B. (2008). The roles of "face" and "non-face" areas during individual face perception: Evidence by fMRI adaptation in a brain-damaged prosopagnosic patient. *NeuroImage*, *40*(1), 318–332. https://doi.org/10.1016/j.neuroimage.2007.11.012

Duncan-Johnson. (1981). Young Psychophysiologist Award address, 1980. P300 latency: a new metric of information processing. *Psychophysiology*, *18*(3), 207–215. https://doi.org/10.1111/psyp.1981.18.issue-3

Duncan-Johnson, C., & Kopell, B. (1981). The Stroop effect: brain potentials localize the source of interference. *Science*, *214*(4523), 938–940. https://doi.org/10.1126/science.7302571

Dunovan, K. E., Tremel, J. J., & Wheeler, M. E. (2014). Prior probability and feature predictability interactively bias perceptual decisions. *Neuropsychologia*, *61*(1), 210–221. https://doi.org/10.1016/j.neuropsychologia.2014.06.024

Eklund, A., Nichols, T. E., & Knutsson, H. (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(28), 7900–7905. https://doi.org/10.1073/pnas.1602413113

Faria, C. de A., Alves, H. V. D., & Charchat-Fichman, H. (2015). The most frequently used tests for assessing executive functions in aging. *Dementia & Neuropsychologia*, *9*(2), 149–155. https://doi.org/10.1590/1980-57642015DN92000009

Forstmann, B. U., Brown, S., Dutilh, G., Neumann, J., & Wagenmakers, E.-J. (2010). The neural substrate of prior information in perceptual decision making: a model-based analysis. *Frontiers in Human Neuroscience*, *4*, 40. https://doi.org/10.3389/fnhum.2010.00040

Forstmann, Ratcliff, R., & Wagenmakers, E.-J. (2016). Sequential Sampling Models in Cognitive Neuroscience: Advantages, Applications, and Extensions. *Annual Review of Psychology*, *67*(1), 641–666. https://doi.org/10.1146/annurev-psych-122414-033645

Francken, J. C., Kok, P., Hagoort, P., & de Lange, F. P. (2015). The behavioral and neural effects of language on motion perception. *Journal of Cognitive Neuroscience*, *27*(1), 175–184. https://doi.org/10.1162/jocn_a_00682

Freeman, J. B., Rule, N. O., Adams Jr, R. B., & Ambady, N. (2010). The Neural Basis of Categorical Face Perception: Graded Representations of Face Gender in Fusiform and Orbitofrontal Cortices. *Cerebral Cortex*, *20*, 1314–1322. https://doi.org/10.1093/cercor/bhp195

Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*. https://doi.org/10.1080/02643290442000310

Gamerman and Lopes, H. F. (2006). Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, 2nd Edn. *London: Taylor and Francis.*, *323*.

Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, *7*(4), 457–472. https://doi.org/10.1214/ss/1177011136

Gennari, S. P., MacDonald, M. C., Postle, B. R., & Seidenberg, M. S. (2007). Context-dependent interpretation of words: Evidence for interactive neural processes. *NeuroImage*, *35*(3), 1278–1286. https://doi.org/10.1016/j.neuroimage.2007.01.015

Gilbert, C. D., & Sigman, M. (2007). Brain States: Top-Down Influences in Sensory Processing. *Neuron*. https://doi.org/10.1016/j.neuron.2007.05.019

Gilbert, Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(2), 489–494. https://doi.org/10.1073/pnas.0509868103

Gilbert, Regier, T., Kay, P., & Ivry, R. B. (2008). Support for lateralization of the Whorf effect beyond the realm of color discrimination. *Brain and Language*, *105*(2), 91–98. https://doi.org/10.1016/j.bandl.2007.06.001

Glaser, W. R., & Düngelhoff, F. J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology. Human Perception and Performance*, *10*(5), 640–654. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6238124

Gold, J. I., & Shadlen, M. N. (2007). The Neural Basis of Decision Making. *Annual Review of Neuroscience*, *30*(1), 535–574. https://doi.org/10.1146/annurev.neuro.29.051605.113038

Gomez, P., Perea, M., & Ratcliff, R. (2013). A diffusion model account of masked versus unmasked priming: Are they qualitatively different? *Journal of Experimental Psychology: Human Perception and Performance*, *39*(6), 1731–1740. https://doi.org/10.1037/a0032333

Grappe, A., Sarma, S. V., Sacré, P., González-Martínez, J., Liégeois-Chauvel, C., & Alario, F.-X. (2018). An intracerebral exploration of functional connectivity during word production. *Journal of Computational Neuroscience*, 1–16. https://doi.org/10.1007/s10827-018-0699-3

Guo, C. C., Gorno-Tempini, M. L., Gesierich, B., Henry, M., Trujillo, A., Shany-Ur, T., … Seeley, W. W. (2013). Anterior temporal lobe degeneration produces widespread network-driven dysfunction. *Brain*, *136*(10), 2979–2991. https://doi.org/10.1093/brain/awt222

Hanes, D. P., & Schall, J. D. (1996). Neural control of voluntary movement initiation. *Science*, *274*(5286), 427–430. https://doi.org/10.1126/science.274.5286.427

Harvey, D. Y., & Schnur, T. T. (2016). Different Loci of Semantic Interference in Picture Naming vs. Word-Picture Matching Tasks. *Frontiers in Psychology*, *7*, 710. https://doi.org/10.3389/fpsyg.2016.00710

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic Representation of Action Words in Human Motor and Premotor Cortex. *Neuron*, *41*(2), 301–307. https://doi.org/10.1016/S0896-6273(03)00838-9

Haxby, J. V, Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science (New York, N.Y.)*, *293*(5539), 2425–2430. https://doi.org/10.1126/science.1063736

Henson. (2003). Neuroimaging studies of priming. *Progress in Neurobiology*, *70*(1), 53–81. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12927334

Hermann, P., Bankó, É. M., Gál, V., & Vidnyánszky, Z. (2015). Neural basis of identity information extraction from noisy face images. *Journal of Neuroscience*, *35*(18), 7165–7173. https://doi.org/10.1523/JNEUROSCI.3572-14.2015

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, *92*(1–2), 67–99. https://doi.org/10.1016/J.COGNITION.2003.10.011

Hillyard, S. A., Squires, K. C., Bauer, J. W., & Lindsay, P. H. (1971). Evoked potential correlates of auditory signal detection. *Science*, *172*(3990), 1357–1360. https://doi.org/10.1126/science.172.3990.1357

Hoffman, P. (2018). An individual differences approach to semantic cognition: Divergent effects of age on representation, retrieval and selection. *Scientific Reports*, *8*(1), 8145. https://doi.org/10.1038/s41598-018-26569-0

Hoffman, P., Lambon Ralph, M. A., & Rogers, T. T. (2013). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods*, *45*(3), 718–730. https://doi.org/10.3758/s13428-012-0278-x

Hoffman, P., McClelland, J. L., & Lambon Ralph, M. A. (2018). Concepts, control, and context: A connectionist account of normal and disordered semantic cognition. *Psychological Review*, *125*(3), 293–328. https://doi.org/10.1037/rev0000094

Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: experimental and computational studies. *Cognition*, *100*(3), 464–482. https://doi.org/10.1016/j.cognition.2005.02.006

Humphreys, G. W., & Forde, E. M. E. (2005). Naming a giraffe but not an animal: Base-level but not superordinate naming in a patient with impaired semantics. *Cognitive Neuropsychology*, *22*(5), 539–558. https://doi.org/10.1080/02643290442000176

JASP Team. (2018). JASP Team (2018). JASP (Version 0.8.6)[Computer software]. Retrieved from https://jasp-stats.org

Jefferies, E. (2013). The neural basis of semantic cognition: Converging evidence from neuropsychology, neuroimaging and TMS. *Cortex*, *49*(3), 611–625. https://doi.org/10.1016/j.cortex.2012.10.008

Jefferies, E., & Lambon Ralph, M. A. (2006). Semantic impairment in stroke aphasia versus semantic dementia: a case-series comparison. *Brain*, *129*(8), 2132–2147. https://doi.org/10.1093/brain/awl153

Jefferies, E., Patterson, K., & Ralph, M. A. L. (2008). Deficits of knowledge versus executive control in semantic cognition: insights from cued naming. *Neuropsychologia*, *46*(2), 649–658. https://doi.org/10.1016/j.neuropsychologia.2007.09.007

Jefferies, E., Rogers, T. T., Hopper, S., & Ralph, M. A. L. (2010). "Pre-semantic" cognition revisited: critical differences between semantic aphasia and semantic dementia. *Neuropsychologia*, *48*(1), 248–261. https://doi.org/10.1016/j.neuropsychologia.2009.09.011

Jehee, J. F. M., Brady, D. K., & Tong, F. (2011). Attention improves encoding of task relevant features in the human visual cortex. *Journal of Neuroscience*, *31*(22), 8210–8219. https://doi.org/10.1523/JNEUROSCI.6153-09.2011

Jenkinson, Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, *17*(2), 825–841. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12377157

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage*, *62*(2), 782–790. https://doi.org/10.1016/j.neuroimage.2011.09.015

Jenkinson, & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, *5*(2), 143–156. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11516708

Kan, I. P., & Thompson-Schill, S. L. (2004). Effect of name agreement on prefrontal activity during overt and covert picture naming. *Cognitive, Affective & Behavioral Neuroscience*, *4*(1), 43–57. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15259888

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: A cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences*. Royal Society. https://doi.org/10.1098/rstb.2006.1934

Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, *90*(430), 773–795. https://doi.org/10.1080/01621459.1995.10476572

Kaul, C., Rees, G., & Ishai, A. (2011). The Gender of Face Stimuli is Represented in Multiple Regions in the Human Brain. *Frontiers in Human Neuroscience*, *4*, 238. https://doi.org/10.3389/fnhum.2010.00238

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, *42*(3), 643–650. https://doi.org/10.3758/BRM.42.3.643

Kiefer, M., & Pulvermüller, F. (2012). Conceptual representations in mind and brain: theoretical developments, current evidence and future directions. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *48*(7), 805–825. https://doi.org/10.1016/j.cortex.2011.04.006

Kinoshita, S., de Wit, B., Aji, M., & Norris, D. (2017). Evidence accumulation in the integrated and primed Stroop tasks. *Memory & Cognition*, *45*(5), 824–836. https://doi.org/10.3758/s13421-017-0701-8

Klauer, & Musch, J. (2003). *The Psychology of Evaluation: Affective Processes in Cognition and Emotion - Google Boeken*. Retrieved from https://books.google.nl/books?hl=nl&lr=&id=t1h6AgAAQBAJ&oi=fnd&pg=PA9&dq=klauer+2003+affective+priming&ots=bBIA87KF4_&sig=rC41jScrG1-KpROOfSwyN-RpY69c#v=onepage&q=klauer 2003 affective priming&f=false

Klepp, A., van Dijk, H., Niccolai, V., Schnitzler, A., & Biermann-Ruben, K. (2019). Action verb processing specifically modulates motor behaviour and sensorimotor neuronal oscillations. *Scientific Reports*, *9*(1), 15985. https://doi.org/10.1038/s41598-019-52426-9

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, *13*(10), 1292–1298. https://doi.org/10.1038/nn.2635

Kuipers, J.-R., & La Heij, W. (2008). Semantic facilitation in category and action naming: Testing the message-congruency account. *Journal of Memory and Language*, *58*(1), 123–139. https://doi.org/10.1016/j.jml.2007.05.005

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: the P300 as a measure of stimulus evaluation time. *Science (New York, N.Y.)*, *197*(4305), 792–795. https://doi.org/10.1126/science.887923

Lambon Ralph, M. A., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, *18*(1), 42–55. https://doi.org/10.1038/nrn.2016.150

Lambon Ralph, M. A., & Patterson, K. (2008). *Generalization and Differentiation in Semantic Memory. Annals of the New York Academy of Sciences*, *1124*(1), 61–76. https://doi.org/10.1196/annals.1440.006

Landau, A. N., Aziz-Zadeh, L., & Ivry, R. B. (2010). The Influence of Language on Perception: Listening to Sentences about Faces Affects the Perception of Faces. *Journal of Neuroscience*, *30*(45), 15254–15261. https://doi.org/10.1523/JNEUROSCI.2046-10.2010

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *The Behavioral and Brain Sciences*, *22*(1), 1–38; discussion 38-75. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11301520

Lewin, C., & Herlitz, A. (2002). Sex differences in face recognition—Women's faces make the difference. *Brain and Cognition*, *50*(1), 121–128. https://doi.org/10.1016/S0278-2626(02)00016-7

Liang, F., Paulo, R., Molina, G., Clyde, M. A., & Berger, J. O. (2008). Mixtures of *g* Priors for Bayesian Variable Selection. *Journal of the American Statistical Association*, *103*(481), 410–423. https://doi.org/10.1198/016214507000001337

Liang, & Luo, Y. (2020). A Comprehensive Comparison of Model Selection Methods for Testing Factorial Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(3), 380–395. https://doi.org/10.1080/10705511.2019.1649983

Lim, S. L., O'Doherty, J. P., & Rangel, A. (2011). The decision value computations in the vmPFC and striatum use a relative value code that is guided by visual attention. *Journal of Neuroscience*, *31*(37), 13214–13223. https://doi.org/10.1523/JNEUROSCI.1246-11.2011

Lisman. (2005). The theta/gamma discrete phase code occuring during the hippocampal phase precession may be a more general brain coding scheme. *Hippocampus*, *15*(7), 913–922. https://doi.org/10.1002/hipo.20121

Lisman, & Idiart, M. A. P. (1995). Storage of 7 ± 2 short-term memories in oscillatory subcycles. *Science*, *267*(5203), 1512–1515. https://doi.org/10.1126/science.7878473

Love, B. C. (2015). The algorithmic level is the bridge between computation and brain. *Topics in Cognitive Science*, *7*(2), 230–242. https://doi.org/10.1111/tops.12131

Lupker, S. J. (1979). The semantic nature of response competition in the picture-word interference task. *Memory & Cognition*, *7*(6), 485–495. https://doi.org/10.3758/bf03198265

Lupyan, G., Spivey, M. J., Frank, M., Wu, L., & Wade, A. (2010). Making the Invisible Visible: Verbal but Not Visual Cues Enhance Visual Detection. *PLoS ONE*, *5*(7), e11452. https://doi.org/10.1371/journal.pone.0011452

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, *9*(11), 1432–1438. https://doi.org/10.1038/nn1790

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory : a user's guide*. Lawrence Erlbaum.

Madden, D. J., Siciliano, R. E., Tallman, C. W., Monge, Z. A., Voss, A., & Cohen, J. R. (2019). Response-level processing during visual feature search: Effects of frontoparietal activation and adult age. *Attention, Perception, & Psychophysics*, *82*(1), 330–349. https://doi.org/10.3758/s13414-019-01823-3

Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). Lexical selection is not by competition: A reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 503–535. https://doi.org/10.1037/0278-7393.33.3.503

Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*, *92*, 57–78. https://doi.org/10.1016/j.jml.2016.04.001

Marr, D. (1982). Vision: A computational investigation into the human representation and processing of visual information. *San Francisco, CA: W.H. Freeman*.

Martin, A. (2007). The Representation of Object Concepts in the Brain. *Annual Review of Psychology*, *58*(1), 25–45. https://doi.org/10.1146/annurev.psych.57.102904.190143

Martins, M. J. D., Krause, C., Neville, D. A., Pino, D., Villringer, A., & Obrig, H. (2019). Recursive hierarchical embedding in vision is impaired by posterior middle temporal gyrus lesions. *Brain*, *142*(10), 3217–3229. https://doi.org/10.1093/brain/awz242

Mazurek, M. E., Roitman, J. D., Ditterich, J., & Shadlen, M. N. (2003). A Role for Neural Integrators in Perceptual Decision Making. *Cerebral Cortex*, *13*(11), 1257–1269. https://doi.org/10.1093/cercor/bhg097

McKelvie, S. J. (1987). Sex differences, lateral reversal, and pose as factors in recognition memory for photographs of Faces. *Journal of General Psychology*, *114*(1), 13–37. https://doi.org/10.1080/00221309.1987.9711052

Mecklinger, A., Kramer, A. F., & Strayer, D. L. (1992). Event Related Potentials and EEG Components in a Semantic Memory Search Task. *Psychophysiology*, *29*(1), 104–119. https://doi.org/10.1111/j.1469-8986.1992.tb02021.x

Meteyard, L., Bahrami, B., & Vigliocco, G. (2007). Motion Detection and Motion Verbs. *Psychological Science*, *18*(11), 1007–1013. https://doi.org/10.1111/j.1467-9280.2007.02016.x

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*(2), 227–234. https://doi.org/10.1037/h0031564

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. In *1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings*. International Conference on Learning Representations, ICLR.

Mirman, D., Landrigan, J.-F., & Britt, A. E. (2017). Taxonomic and thematic semantic systems. *Psychological Bulletin*, *143*(5), 499–520. https://doi.org/10.1037/bul0000092

Mitterer, H., Horschig, J. M., Müsseler, J., & Majid, A. (2009). The Influence of Memory on Perception: It's Not What Things Look Like, It's What You Call Them. *Journal of Experimental Psychology: Learning Memory and Cognition*, *35*(6), 1557–1562. https://doi.org/10.1037/a0017019

Monin, B. (2003). The Warm Glow Heuristic: When Liking Leads to Familiarity. *Journal of Personality and Social Psychology*, *85*(6), 1035–1048. https://doi.org/10.1037/0022-3514.85.6.1035

Moody, T. D., Sasaki, M. A., Bohon, C., Strober, M. A., Bookheimer, S. Y., Sheen, C. L., & Feusner, J. D. (2015). Functional connectivity for face processing in individuals with body dysmorphic disorder and anorexia nervosa. *Psychological Medicine*, *45*(16), 3491–3503. https://doi.org/10.1017/S0033291715001397

Mooney, C. M. (1957). Age in the development of closure ability in children. *Canadian Journal of Psychology*, *11*(4), 219–226. https://doi.org/10.1037/h0083717

Mulder, M. J., Wagenmakers, E. J., Ratcliff, R., Boekel, W., & Forstmann, B. U. (2012). Bias in the brain: A diffusion model analysis of prior probability and potential payoff. *Journal of Neuroscience*, *32*(7), 2335–2343. https://doi.org/10.1523/JNEUROSCI.4156-11.2012

Mulder, van Maanen, L., & Forstmann, B. U. (2014). Perceptual decision neurosciences – A model-based review. *Neuroscience*, *277*, 872–884. https://doi.org/10.1016/j.neuroscience.2014.07.031

Nagel, I. E., Schumacher, E. H., Goebel, R., & D'Esposito, M. (2008). Functional MRI investigation of verbal selection mechanisms in lateral prefrontal cortex. *NeuroImage*, *43*(4), 801–807. https://doi.org/10.1016/j.neuroimage.2008.07.017

Navarrete, E., Mahon, B. Z., & Caramazza, A. (2010). The cumulative semantic cost does not reflect lexical selection by competition. *Acta Psychologica*, *134*(3), 279–289. https://doi.org/10.1016/j.actpsy.2010.02.009

Navarro, D. J., & Fuss, I. G. (2009). Fast and accurate calculations for first-passage times in Wiener diffusion models. *Journal of Mathematical Psychology*. Academic Press Inc. https://doi.org/10.1016/j.jmp.2009.02.003

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In Besner D. & Humphreys G. W. (Eds.), *Basic processes in reading: visual word recognition: a selective review of current finding and theories*. (pp. 264–336). Hillsdale, NJ: Erlbaum. Retrieved from http://garfield.library.upenn.edu/classics1991/A1991FL03200001.pdf

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers : A Journal of the Psychonomic Society, Inc*, *36*(3), 402–407. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15641430

Noonan, K. A., Jefferies, E., Corbett, F., & Lambon Ralph, M. A. (2010). Elucidating the nature of deregulated semantic cognition in semantic aphasia: Evidence for the roles of prefrontal and temporo-parietal cortices. *Journal of Cognitive Neuroscience*, *22*(7), 1597–1613. https://doi.org/10.1162/jocn.2009.21289

Noonan, K. A., Jefferies, E., Visser, M., & Lambon Ralph, M. A. (2013). Going beyond Inferior Prefrontal Involvement in Semantic Control: Evidence for the Additional Contribution of Dorsal Angular Gyrus and Posterior Middle Temporal Cortex. *Journal of Cognitive Neuroscience*, *25*(11), 1824–1850. https://doi.org/10.1162/jocn_a_00442

Noppeney, Phillips, J., & Price, C. (2004). The neural areas that control the retrieval and selection of semantics. *Neuropsychologia*, *42*(9), 1269–1280. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2003.12.014

Noppeney, U., Ostwald, D., & Werner, S. (2010). Perceptual decisions formed by accumulation of audiovisual evidence in prefrontal cortex. *Journal of Neuroscience*, *30*(21), 7434–7446. https://doi.org/10.1523/JNEUROSCI.0455-10.2010

Nozari, N. (2017). The dual origin of semantic errors in access deficits: activation deficit vs. inhibition deficit. *Frontiers in Human Neuroscience*, *11*. https://doi.org/10.3389/conf.fnhum.2017.223.00100

Nozari, N., & Hepner, C. R. (2018). To select or to wait? The importance of criterion setting in debates of competitive lexical selection. *Cognitive Neuropsychology*, 1–15. https://doi.org/10.1080/02643294.2018.1476335

Nunez, M. D., Gosai, A., Vandekerckhove, J., & Srinivasan, R. (2019). The latency of a visual evoked potential tracks the onset of decision making. *NeuroImage*, *197*, 93–108. https://doi.org/10.1016/j.neuroimage.2019.04.052

Oishi, K., Toma, K., Bagarinao, E. T., Matsuo, K., Nakai, T., Chihara, K., & Fukuyama, H. (2005). Activation of the precuneus is related to reduced reaction time in serial reaction time tasks. *Neuroscience Research*, *52*(1), 37–45. https://doi.org/10.1016/j.neures.2005.01.008

Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2007). Cumulative semantic interference as learning. *Brain and Language*, *103*(1–2), 175–176. https://doi.org/10.1016/j.bandl.2007.07.102

Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. *Cognition*, *114*(2), 227–252. https://doi.org/10.1016/j.cognition.2009.09.007

Oppenheimer, D. M., & Frank, M. C. (2008). A rose in any other font would not smell as sweet: Effects of perceptual fluency on categorization. *Cognition*, *106*(3), 1178–1194. https://doi.org/10.1016/j.cognition.2007.05.010

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. *Nature Reviews Neuroscience*, *8*(12), 976–987. https://doi.org/10.1038/nrn2277

Peebles, D., & Cooper, R. P. (2015). Thirty years after Marr's vision: Levels of analysis in cognitive science. *Topics in Cognitive Science*, *7*(2), 187–190. https://doi.org/10.1111/tops.12137

Philiastides, M. G., & Sajda, P. (2006). Temporal Characterization of the Neural Correlates of Perceptual Decision Making in the Human Brain. *Cerebral Cortex*, *16*(4), 509–518. https://doi.org/10.1093/cercor/bhi130

Piai, V., & Knight, R. T. (2018). Lexical selection with competing distractors: Evidence from left temporal lobe lesions. *Psychonomic Bulletin & Review*, *25*(2), 710–717. https://doi.org/10.3758/s13423-017-1301-0

Piai, V., Ries, S. K., & Swick, D. (2016). Lesions to Lateral Prefrontal Cortex Impair Lexical Interference Control in Word Production. *Frontiers in Human Neuroscience*, *9*, 721. https://doi.org/10.3389/fnhum.2015.00721

Piai, V., Roelofs, A., Acheson, D. J., & Takashima, A. (2013a). Attention for speaking: domain-general control from the anterior cingulate cortex in spoken word production. *Frontiers in Human Neuroscience*, *7*, 832. https://doi.org/10.3389/fnhum.2013.00832

Piai, V., Roelofs, A., Jensen, O., Schoffelen, J.-M., & Bonnefond, M. (2014). Distinct Patterns of Brain Activity Characterise Lexical Activation and Competition in Spoken Word Production. *PLoS ONE*, *9*(2), e88674. https://doi.org/10.1371/journal.pone.0088674

Potter, K. W., Donkin, C., & Huber, D. E. (2018). The elimination of positive priming with increasing prime duration reflects a transition from perceptual fluency to disfluency rather than bias against primed words. *Cognitive Psychology*, *101*, 1–28. https://doi.org/10.1016/j.cogpsych.2017.11.004

Puri, A. M., Wojciulik, E., & Ranganath, C. (2009). Category expectation modulates baseline and stimulus-evoked activity in human inferotemporal cortex. *Brain Research*, *1301*, 89–99. https://doi.org/10.1016/j.brainres.2009.08.085

Python, G., Glize, B., & Laganaro, M. (2018). The involvement of left inferior frontal and middle temporal cortices in word production unveiled by greater facilitation effects following brain damage. *Neuropsychologia*, *121*, 122–134. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2018.10.026

Raftopoulos, A., & Lupyan, G. (2018). Editorial: Pre-cueing Effects on Perception and Cognitive Penetrability. *Frontiers in Psychology*, *9*(FEB), 230. https://doi.org/10.3389/fpsyg.2018.00230

Ralph, M. A. L., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, *18*(1), 42–55. https://doi.org/10.1038/nrn.2016.150

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108. https://doi.org/10.1037/0033-295X.85.2.59

Ratcliff, R., Cherian, A., & Segraves, M. (2003). A comparison of Macaque behavior and superior colliculus neuronal activity to predictions from models of two-choice decisions. *Journal of Neurophysiology*, *90*(3), 1392–1407. https://doi.org/10.1152/jn.01049.2002

Ratcliff, R., & Childers, R. (2015). Individual differences and fitting methods for the two-choice diffusion model of decision making. *Decision*, *2*(4), 237–279. https://doi.org/10.1037/dec0000030

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A Diffusion Model Account of the Lexical Decision Task. *Psychological Review*, *111*(1), 159–182. https://doi.org/10.1037/0033-295X.111.1.159

Ratcliff, R., & McKoon, G. (2008a). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Computation*, *20*(4), 873–922. https://doi.org/10.1162/neco.2008.12-06-420

Ratcliff, R., McKoon, G., Hintzman, D., Mandler, G., Masson, M., & Jo Nissen, M. (1988). A Retrieval Theory of Priming in Memory, *95*(3), 385–408. Retrieved from http://star.psy.ohio-state.edu/wp/pdf/Papers/psychrev88b.pdf

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion Decision Model: Current Issues and History. *Trends in Cognitive Sciences*, *20*(4), 260–281. https://doi.org/10.1016/j.tics.2016.01.007

Ratcliff, R., Thapar, A., Gomez, P., & McKoon, G. (2004). A diffusion model analysis of the effects of aging in the lexical-decision task. *Psychology and Aging*, *19*(2), 278–289. https://doi.org/10.1037/0882-7974.19.2.278

Ries, S. K., Greenhouse, I., Dronkers, N. F., Haaland, K. Y., & Knight, R. T. (2014). Double dissociation of the roles of the left and right prefrontal cortices in anticipatory regulation of action. *Neuropsychologia*, *63*, 215–225. https://doi.org/10.1016/j.neuropsychologia.2014.08.026

Ries, S. K., Karzmark, C. R., Navarrete, E., Knight, R. T., & Dronkers, N. F. (2015). Specifying the role of the left prefrontal cortex in word selection. *Brain and Language*, *149*, 135–147. https://doi.org/10.1016/J.BANDL.2015.07.007

Rissman, J., Eliassen, J. C., & Blumstein, S. E. (2003). An Event-Related fMRI Investigation of Implicit Semantic Priming. *Journal of Cognitive Neuroscience*, *15*(8), 1160–1175. https://doi.org/10.1162/089892903322598120

Roelofs. (2018). A unified computational account of cumulative semantic, semantic blocking, and semantic distractor effects in picture naming. *Cognition*, *172*, 59–72. https://doi.org/10.1016/J.COGNITION.2017.12.007

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*(1–3), 107–142. https://doi.org/10.1016/0010-0277(92)90041-F

Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., & Patterson, K. (2004). Structure and Deterioration of Semantic Memory: A Neuropsychological and Computational Investigation. *Psychological Review*, *111*(1), 205–235. https://doi.org/10.1037/0033-295X.111.1.205

Rorden, C., & Brett, M. (2000). Stereotaxic display of brain lesions. *Behavioural Neurology*, *12*(4), 191–200. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11568431

Rossell, S. L., Price, C. J., & Nobre, A. C. (2003). The anatomy and time course of semantic priming investigated by fMRI and ERPs. *Neuropsychologia*, *41*(5), 550–564. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12638581

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*, 356–374. https://doi.org/10.1016/j.jmp.2012.08.001

Ruff, I., Blumstein, S. E., Myers, E. B., & Hutchison, E. (2008). Recruitment of anterior and posterior structures in lexical-semantic processing: an fMRI study comparing implicit and explicit tasks. *Brain and Language*, *105*(1), 41–49. https://doi.org/10.1016/j.bandl.2008.01.003

Samaha, J., Boutonnet, B., Postle, B. R., & Lupyan, G. (2018). Effects of meaningfulness on perception: Alpha-band oscillations carry perceptual expectations and influence early visual responses. *Scientific Reports*, *8*(1), 6606. https://doi.org/10.1038/s41598-018-25093-5

Schendan, H. E., & Kutas, M. (2007). Neurophysiological evidence for the time course of activation of global shape, part, and local contour representations during visual object categorization and memory. *Journal of Cognitive Neuroscience*, *19*(5), 734–749. https://doi.org/10.1162/jocn.2007.19.5.734

Schnur, T. T., Schwartz, M. F., Kimberg, D. Y., Hirshorn, E., Coslett, H. B., & Thompson-Schill, S. L. (2009). Localizing interference during naming: Convergent neuroimaging and neuropsychological evidence for the function of Broca’s area. *Proceedings of the National Academy of Sciences*, *106*(1), 322–327. https://doi.org/10.1073/pnas.0805874106

Schnur, Tatiana T., Schwartz, M. F., Brecher, A., & Hodgson, C. (2006). Semantic interference during blocked-cyclic naming: Evidence from aphasia. *Journal of Memory and Language*, *54*(2), 199–227. https://doi.org/10.1016/J.JML.2005.10.002

Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences*, *32*(1), 9–18. https://doi.org/10.1016/j.tins.2008.09.012

Schwartz, M. F., Kimberg, D. Y., Walker, G. M., Faseyitan, O., Brecher, A., Dell, G. S., & Coslett, H. B. (2009). Anterior temporal involvement in semantic word retrieval: voxel-based lesion-symptom mapping evidence from aphasia. *Brain : A Journal of Neurology*, *132*(Pt 12), 3411–3427. https://doi.org/10.1093/brain/awp284

Schwarz. (2004). Metacognitive experiences in consumer judgment and decision making. *Journal of Consumer Psychology*, *14*(4), 332–348. https://doi.org/10.1207/s15327663jcp1404_2

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Shadlen, M. N., & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology*, *86*(4), 1916–1936. https://doi.org/10.1152/jn.2001.86.4.1916

Simanova, I., Francken, J. C., de Lange, F. P., & Bekkering, H. (2016). Linguistic priors shape categorical perception. *Language, Cognition and Neuroscience*, *31*(1), 159–165. https://doi.org/10.1080/23273798.2015.1072638

Sitnikova, T., Holcomb, P. J., Kiyonaga, K. A., & Kuperberg, G. R. (2008). Two neurocognitive mechanisms of semantic integration during the comprehension of visual real-world events. *Journal of Cognitive Neuroscience*, *20*(11), 2037–2057. https://doi.org/10.1162/jocn.2008.20143

Smith, & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences*, *27*(3), 161–168. https://doi.org/10.1016/j.tins.2004.01.006

Smith, Ratcliff, R., & Wolfgang, B. J. (2004). Attention orienting and the time course of perceptual decisions: Response time distributions with masked and unmasked displays. In *Vision Research* (Vol. 44, pp. 1297–1320). Pergamon. https://doi.org/10.1016/j.visres.2004.01.002

Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, *17*(3), 143–155. https://doi.org/10.1002/hbm.10062

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., … Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, *23 Suppl 1*, S208-19. https://doi.org/10.1016/j.neuroimage.2004.07.051

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *64*(4), 583–616. https://doi.org/10.1111/1467-9868.00353

Spitzer, M., Bellemann, M. E., Kammer, T., Gückel, F., Kischka, U., Maier, S., … Brix, G. (1996). Functional MR imaging of semantic information processing and learning-related effects using psychometrically controlled stimulation paradigms. *Cognitive Brain Research*, *4*(3), 149–161. https://doi.org/10.1016/S0926-6410(96)00028-6

Squires, K. C., Hillyard, S. A., & Lindsay, P. H. (1973). Vertex potentials evoked during auditory signal detection: Relation to decision criteria. *Perception & Psychophysics*, *14*(2), 265–272. https://doi.org/10.3758/BF03212388

Tagliabue, C. F., Veniero, D., Benwell, C. S. Y., Cecere, R., Savazzi, S., & Thut, G. (2019). The EEG signature of sensory evidence accumulation during decision formation closely tracks subjective perceptual experience. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-41024-4

Tan, Chan, A. H. D., Kay, P., Khong, P.-L., Yip, L. K. C., & Luke, K.-K. (2008). Language affects patterns of brain activation associated with perceptual decision. *Proceedings of the National Academy of Sciences*, *105*(10), 4004–4009. https://doi.org/10.1073/pnas.0800055105

Thompson-Schill, D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(26), 14792–14797. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9405692

Thompson-Schill, S. L., Jonides, J., Marshuetz, C., Smith, E. E., D'Esposito, M., Kan, I. P., … Swick, D. (2002). Effects of frontal lobe damage on interference effects in working memory. *Cognitive, Affective & Behavioral Neuroscience*, *2*(2), 109–120. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12455679

Todorova, L., & Neville, D. A. (2020). Associative and Identity Words Promote the Speed of Visual Categorization: A Hierarchical Drift Diffusion Account. *Frontiers in Psychology*, *11*, 955. https://doi.org/10.3389/fpsyg.2020.00955

Tremel, J. J., & Wheeler, M. E. (2015). Content-specific evidence accumulation in inferior temporal cortex during perceptual decision-making. *NeuroImage*, *109*, 35–49. https://doi.org/10.1016/j.neuroimage.2014.12.072

Twomey, D. M., Murphy, P. R., Kelly, S. P., & O'Connell, R. G. (2015). The classic P300 encodes a build-to-threshold decision variable. *European Journal of Neuroscience*, *42*(1), 1636–1643. https://doi.org/10.1111/ejn.12936

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological Review*, *108*(3), 550–592. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11488378

van Bergen, R. S., Ma, W. J., Pratte, M. S., & Jehee, J. F. M. (2015). Sensory uncertainty decoded from visual cortex predicts behavior. *Nature Neuroscience*, *18*(12), 1728–1730. https://doi.org/10.1038/nn.4150

Voss, A., Nagler, M., & Lerche, V. (2013). Diffusion Models in Experimental Psychology. *Experimental Psychology*, *60*(6), 385–402. https://doi.org/10.1027/1618-3169/a000218

Voss, A., Rothermund, K., Gast, A., & Wentura, D. (2013). Cognitive processes in associative and categorical priming: A diffusion model analysis. *Journal of Experimental Psychology: General*, *142*(2), 536–559. https://doi.org/10.1037/a0029459

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: an empirical validation. *Memory & Cognition*, *32*(7), 1206–1220. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15813501

Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., … Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, *25*(1), 35–57. https://doi.org/10.3758/s13423-017-1343-3

Wagenmakers, Van Der Maas, H. L. J., & Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin and Review*, *14*(1), 3–22. https://doi.org/10.3758/BF03194023

Wagner, A. D., Paré-Blagoev, E. J., Clark, J., & Poldrack, R. A. (2001). Recovering meaning: left prefrontal cortex guides controlled semantic retrieval. *Neuron*, *31*(2), 329–338. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11502262

Warrington, E. K., & Shallice, T. (1984). Category specific semantic impairments. *Brain : A Journal of Neurology*, *107 (Pt 3)*, 829–854. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6206910

Wentura, D. (2000). Dissociative affective and associative priming effects in the lexical decision task: yes versus no responses to word targets reveal evaluative judgment tendencies. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *26*(2), 456–469. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10764106

Wentura, D., & Degner, J. (2010). A practical guide to sequential priming and related tasks. In Gawronski B. & Payne B.K. (Ed.), *Handbook of implicit social cognition: Measurement, theory and applications* (Guilford Press, pp. 95–115). New York: Guilford Press. Retrieved from http://dare.uva.nl/search?arno.record.id=375383

White, C. N., & Poldrack, R. A. (2014). Decomposing bias in different types of simple decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(2), 385–398. https://doi.org/10.1037/a0034851

Whitney, C., Kirk, M., O'Sullivan, J., Lambon Ralph, M. A., & Jefferies, E. (2011). The Neural Organization of Semantic Control: TMS Evidence for a Distributed Network in Left Inferior Frontal and Posterior Middle Temporal Gyrus. *Cerebral Cortex*, *21*(5), 1066–1075. https://doi.org/10.1093/cercor/bhq180

Whitney, Carin, Kirk, M., O'Sullivan, J., Lambon Ralph, M. A., & Jefferies, E. (2012). Executive Semantic Processing Is Underpinned by a Large-scale Neural Network: Revealing the Contribution of Left Prefrontal, Posterior Temporal, and Parietal Cortex to Controlled Retrieval and Selection Using TMS. *Journal of Cognitive Neuroscience*, *24*(1), 133–147. https://doi.org/10.1162/jocn_a_00123

Whittlesea, B. W. A., & Williams, L. D. (1998). Why do strangers feel familiar, but friends don't? A discrepancy-attribution account of feelings of familiarity. *Acta Psychologica*, *98*(2–3), 141–165. https://doi.org/10.1016/s0001-6918(97)00040-1

Whittlesea, B. W. A., & Williams, L. D. (2000). The Source of Feelings of Familiarity: The Discrepancy-Attribution Hypothesis. *Journal of Experimental Psychology: Learning Memory and Cognition*, *26*(3), 547–565. https://doi.org/10.1037/0278-7393.26.3.547

Whorf, B. L., & Carroll, J. B. (1956). *Language, thought, and reality : selected writings*. Technology Press of Massachusetts Institute of Technology.

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Frontiers in Neuroinformatics*, *7*, 14. https://doi.org/10.3389/fninf.2013.00014

Wiese, H., Kloth, N., Güllmar, D., Reichenbach, J. R., & Schweinberger, S. R. (2012). Perceiving age and gender in unfamiliar faces: An fMRI study on face categorization. *Brain and Cognition*, *78*(2), 163–168. https://doi.org/10.1016/j.bandc.2011.10.012

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: the SHINE toolbox. *Behavior Research Methods*, *42*(3), 671–684. https://doi.org/10.3758/BRM.42.3.671

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(19), 7780–7785. https://doi.org/10.1073/pnas.0701644104

Woolrich, M. W., Behrens, T. E. J., Beckmann, C. F., Jenkinson, M., & Smith, S. M. (2004). Multilevel linear modelling for FMRI group analysis using Bayesian inference. *NeuroImage*, *21*(4), 1732–1747. https://doi.org/10.1016/J.NEUROIMAGE.2003.12.023

Woolrich, M. W., Behrens, T. E. J., & Smith, S. M. (2004). Constrained linear basis sets for HRF modelling using Variational Bayes. *NeuroImage*, *21*(4), 1748–1761. https://doi.org/10.1016/J.NEUROIMAGE.2003.12.024

Woolrich, M. W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., … Smith, S. M. (2009). Bayesian analysis of neuroimaging data in FSL. *NeuroImage*, *45*(1), S173–S186. https://doi.org/10.1016/j.neuroimage.2008.10.055

Woolrich, M. W., Ripley, B. D., Brady, M., & Smith, S. M. (2001). Temporal Autocorrelation in Univariate Linear Modeling of FMRI Data. *NeuroImage*, *14*(6), 1370–1386. https://doi.org/10.1006/NIMG.2001.0931

Wright, D. B., & Sladden, B. (2003). An own gender bias and the importance of hair in face recognition. *Acta Psychologica*, *114*(1), 101–114. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12927345

Yum, Y. N., Holcomb, P. J., & Grainger, J. (2011). Words and pictures: An electrophysiological investigation of domain specific processing in native Chinese and English speakers. *Neuropsychologia*, *49*(7), 1910–1922. https://doi.org/10.1016/j.neuropsychologia.2011.03.018

Zhang, J., Rittman, T., Nombela, C., Fois, A., Coyle-Gilchrist, I., Barker, R. A., … Rowe, J. B. (2016). Different decision deficits impair response inhibition in progressive supranuclear palsy and Parkinson's disease. https://doi.org/10.1093/brain/awv331

Zhang, J., & Rowe, J. B. (2014). Dissociable mechanisms of speed-accuracy tradeoff during visual perceptual learning are revealed by a hierarchical drift-diffusion model. *Frontiers in Neuroscience*, *8*(8 APR), 69. https://doi.org/10.3389/fnins.2014.00069

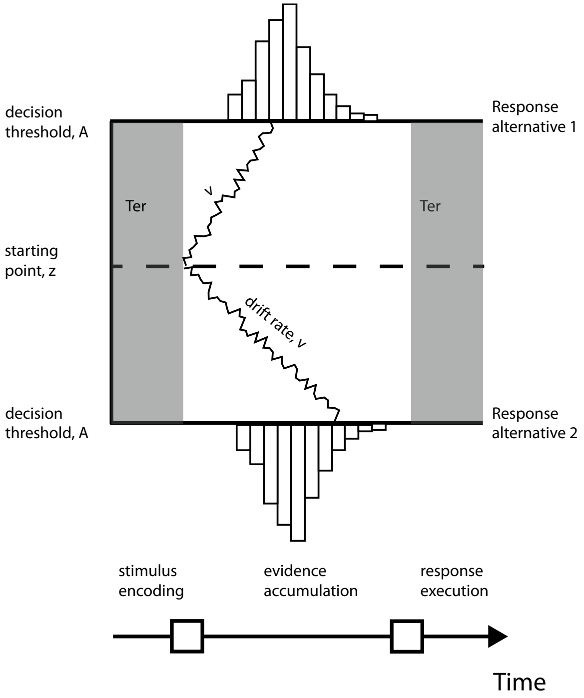
Nederlandse samenvatting

In het dagelijks leven nemen we honderden snelle beslissingen, waar we niet veel over nadenken. Wanneer we bijvoorbeeld een nieuwe collega ontmoeten, merken we automatisch zijn/haar leeftijd en geslacht op zonder er veel aandacht aan te besteden. Deze snelle beslissingen zijn *bottom-up*: ze zijn gebaseerd op de kenmerken van het gezicht en het lichaam van de persoon met wie we praten. Niet alle beslissingen worden bottom-up genomen. Ze kunnen ook *top-down* zijn, dus gebaseerd op de context. Als we bijvoorbeeld een specifieke stem horen die van om de hoek komt, weten we vaak al of het om een mannenstem of een vrouwenstem gaat. Woorden kunnen ook bottom-up worden gebruikt, met andere woorden, ze kunnen een context introduceren die onze perceptie kan beïnvloeden. In experimentele omgevingen zullen deelnemers eerder antwoorden dat ze een huis zien in een dubbelzinnig beeld, als het voorafgegaan wordt door woorden die betrekking hebben op een concept van een huis («schoorsteen», «raam»). In die zin kunnen woorden onze besluitvorming beïnvloeden.

Het effect van woorden op perceptie en besluitvorming is eerder goed gedocumenteerd (zie voor een review Anderson, Chiu, Huette, & Spivey, 2011; Lupyan, 2012). De mechanismen van deze top-down-invloed zijn echter nog steeds niet duidelijk. Er is bijvoorbeeld aangetoond dat deelnemers verschillende strategieën kunnen toepassen bij het integreren van taalkundige en visuele informatie. Hoewel er enig bewijs is dat de perceptuele gevoeligheid van de deelnemers een rol speelt (Meteyard et al., 2007), is gesuggereerd dat er ook een alternatief mechanisme betrokken kan zijn, dat bestaat uit het aanpassen van de beslissingsdrempel (Francken et al., 2015). In dit proefschrift heb ik onderzocht hoe onze besluitvorming wordt beïnvloed door woorden: wat zijn *de mechanismen* van taalbias bij besluitvorming? Wat zijn *neurale onderbouwingen* van taalbias?

Om de mechanismen van taalbias bij besluitvorming te onderzoeken, heb ik een wiskundig framework gebruikt, genaamd Drift Diffusion Model (DDM). Dit biologisch plausibele model houdt in dat elke binaire beslissing («Is het een man?» - «Ja» / «Nee») kan worden beschreven als een proces van bewijsophoping tot een van de beslissingsgrenzen (bovengrens - vrouwelijk antwoord, en ondergrens - mannelijk antwoord, bijvoorbeeld). DDM gebruikt een combinatie van reactietijden en prestatie-nauwkeurigheden die zijn afgeleid van het experiment om de modelparameters af te leiden die verschillende stadia van het besluitvormingsproces beschrijven. Menselijke prestaties op elke binaire taak kunnen worden beschreven aan de hand van de volgende parameters (een ervan of een combinatie ervan, zie **Figuur 1, A**): driftsnelheid (V) verwijst naar de snelheid waarmee bewijsmateriaal wordt verzameld.

Startpunt (z) verwijst naar de informatie over de beslissing die begon voordat de bewijsverzameling werd geïnitieerd. De beslissingsdrempel (A) staat voor de hoeveelheid informatie die nodig is om tot een beslissing te komen, en de niet-beslissingscomponent (Ter) beschrijft de processen die niet direct verband houden met de vorming van de beslissing (motorische uitvoering / snelheid van stimuluscodering).



Figuur 1: (A) het driftverspreidingsmodel met de vier parameters: driftsnelheid (v), beslissingsdrempel (A), startpunt (z), niet-beslissingscomponent (Ter). **(B)** Verwerkingsstadia die worden gekenmerkt door het DDM-model. Verschillende stadia van verwerking worden gemarkeerd door de schaduw in paneel A. Deze geeft aan welke van de DDM-componenten bij welke verwerkingsfasen horen.

Dit model is nuttig om vast te stellen of taal menselijk gedrag beïnvloedt op het niveau van besluitvorming (selectie van beslissingsstrategie) of op het niveau van de responsuitvoering / snelheid van stimuluscodering, wat geen deel uitmaakt van het besluitvormingsproces zelf (zie **Figuur 1, B**). Bovendien maakt dit model het mogelijk om onderscheid te maken tussen twee verschillende beslissingsstrategieën die mogelijk door taal worden beïnvloed: aanpassing van de beslissingsdrempel versus aanpassing van de snelheid van het verzamelen van bewijsmateriaal.

Verder heb ik onderzocht of door taal beïnvloede beslissingsprocessen (zoals vastgesteld door DDM) kunnen worden weerspiegeld in de neurale data. Hiervoor heb ik neuroimaging-technieken gebruikt zoals functionele magnetische resonantie beeldvorming (fMRI) en magneto-encefalografie (MEG). Hoewel fMRI nuttig is voor het verhelderen van de hersenstructuren die kunnen worden geassocieerd met de besluitvormingsprocessen die door taal worden beïnvloed, is MEG met zijn verfijnde temporele resolutie buitengewoon nuttig om te bepalen wanneer de taal onze besluitvorming beïnvloedt.

Het proefschrift bestaat uit vier empirische hoofdstukken waarin ik de mechanismen van taalbias heb onderzocht met een combinatie van wiskundige en neuroimaging-methoden.

In **Hoofdstuk 2** heb ik een respons-priming-paradigma gebruikt, waarbij een kort gepresenteerd woord (gerelateerd aan het concept van een vrouw of een man) werd gevolgd door een afbeelding van een gezicht dat als vrouwelijk of mannelijk moest worden beoordeeld. Woorden kunnen van twee soorten zijn: ofwel associaties ('stropdas') of identiteitswoorden ('man'). Ik toonde aan dat zowel associaties als identiteitswoorden leidden tot een vergemakkelijking van de geslachtscategorisatie van gezichten (d.w.z. het beantwoorden van 'man' of 'vrouw' op gezichten). Dit effect werd in kaart gebracht op zowel de verwerkingssnelheid als de snelheid van de motorische respons: wanneer ze verband hielden met het doel ('man' of 'vrouw'), resulteerden beide primaire typen in een verhoogde driftsnelheid en een snellere niet-beslissingscomponent. Deze resultaten dienen als een bewijs van het principe dat woorden de perceptuele gevoeligheid beïnvloeden in plaats van de aanpassing van de beslissingsdrempel.

In **Hoofdstuk 3**, waarin ik fMRI gebruik, heb ik laten zien dat de snelheid van visuele categorisatie taakafhankelijk is. Bij elke taak werd het woord (vrouwelijk of mannelijk gerelateerd) gevolgd door een gezicht dat moest worden gecategoriseerd als mannelijk of vrouwelijk. Deze twee taken vereisten een verschillende toewijzing van aandacht aan het woord: in de ene taak moesten de deelnemers het woord negeren en het gezicht classificeren (classificatietaak), terwijl ze in de andere taak (verificatietaak) een expliciete overeenkomst moesten maken tussen het woord en het doelwit (dat wil zeggen: antwoord 'ja' als er een match was tussen woord en gezicht). In de classificatietaak resulteerden woorden die overeenkwamen met het doelwit (woord: "man", doelwit: mannelijk gezicht) in een verhoogde driftsnelheid in vergelijking met incongruente woorden (woord: "man", doelwit: vrouwelijk gezicht). In de verificatietaak resulteerden vrouwgerelateerde woorden in een verhoogde driftsnelheid, ongeacht het geslacht van het doelwit, wat suggereert dat de classificatie- en de verificatietaken verschillende prestatiestrategieën vereisen wat betreft het verzamelen van bewijslast om een beslissing te kunnen nemen over een visueel gepresenteerde categorie. Neurale gegevens in de classificatietaak lieten geen verhelderend bewijs zien over de relatie tussen bewijsaccumulatie en het neurale substraat ervan. In de verificatietaak ontdekten we dat de driftsnelheid tijdens incongruente versus congruente primaire woord-gezicht-paren met mannelijke gezichten correleerde met activiteit in de precuneale cortex op het niveau van de gehele hersenen, wat mogelijk de prestatiestrategie tijdens de taak zou kunnen weerspiegelen. Deze studie biedt een nieuw perspectief op het onderwerp van taalperceptie-interacties met behulp van een combinatie van modelgebaseerde en neuroimaging-benaderingen.

In **Hoofdstuk 4**, waarin ik MEG gebruikte, heb ik laten zien dat woorden helpen bij het beslissen over dubbelzinnige afbeeldingen door de snelheid van visuele categorisatie te verhogen. Dat wil zeggen dat proefpersonen sneller reageerden met ‘gezicht’ of ‘huis’ op een dubbelzinnige afbeelding waarop afbeeldingen van zowel een gezicht als een huis elkaar overlapten (elke dubbelzinnige foto werd voorafgegaan door een woord dat verband hield met het concept van gezicht - “wenkbrauw”, of huis -“venster”). Op hersenniveau beïnvloedden woorden de categorisatie ongeveer 300ms na de presentatie van de afbeelding, wat suggereerde dat woorden ambigue visuele waarneming beïnvloeden door aandacht of beslissing te moduleren in plaats van vroege visuele of lexico-semantische processen.

In **Hoofdstuk 5** heb ik met behulp van een laesiesymptoombenadering laten zien dat visuele beslissingen bij mensen met hersenbeschadiging in temporale versus frontale gebieden op een vergelijkbare manier beïnvloed worden, maar fundamenteel anders dan bij neurotypische controles. In dit hoofdstuk heb ik een woord-afbeelding-matchingtaak gebruikt, waarbij deelnemers ”ja” of “nee” moesten antwoorden op een overeenkomst tussen een woord en een afbeelding. Terwijl de taakuitvoering bij controles werd verklaard door de hoeveelheid informatie die nodig was om een beslissing te nemen (beslissingsdrempel), vertoonden personen met frontale of temporale laesies een vergelijkbare disfunctionele modulatie van zowel de driftsnelheid als de beslissingsdrempel. Deze resultaten ondersteunen een diverse maar interactieve rol van lexicaal-semantisch geheugen en semantische controle-mechanismen.

Concluderend, in dit proefschrift heb ik visuo-semantische interacties geformaliseerd vanuit een besluitvormingsperspectief, d.w.z. als een proces van bewijsaccumulatie tot aan de beslissingsdrempel. In het bijzonder heb ik laten zien dat woorden een specifieke strategie van besluitvorming beïnvloeden: perceptuele gevoeligheid in plaats van aanpassing van de beslissingsdrempel. Ondanks het feit dat ik geen correlaties heb gevonden tussen neurale data en modelleringsdata, heb ik richtingen geschetst voor verdere analyses, wat de resultaten en het verhaal van dit proefschrift informatief maakt voor toekomstige studies.

Acknowledgments

I'm on a plane from Amsterdam to Istanbul, which is my Postdoc destination for the upcoming year or so. I decided it is time to reflect on the past 6 years in the Netherlands, during which I did my PhD in Nijmegen.

I was lucky to be selected for the PhD position in the Language in Interaction Consortium, the fruit of Peter Hagoorts's ambitious idea to advance the field of language science via interdisciplinary contributions to the field that otherwise would not be possible. It was a special place to be, with so much scientific freedom and endless possibilities for collaboration. The Dutch academic culture, with its excellent organization and intense avalanche of meetings/lectures, stimulates young minds. In spite of the fact that some will forever be buried under this avalanche, I will always be mesmerized by how inventive and resilient people are and how they are able to create structure from the chaos. Bumpy and rocky, my PhD experience taught me how to be independent and at the end of the journey it empowered me to go further into academic woods. I am truly grateful to have been able to experience this path. So here it goes, the thank-you's.

To Harold and Irina, thank you for hiring me in the first place for this position. Even if it didn't work out between us, I'm grateful for the opportunity.

Peter Hagoort, you have been receptive to new ideas. You supported my projects and made time to read my texts, which I am grateful for. The fact that you took the responsibility for a fell-out-of-the-wagon student meant a lot to me. Thank you for agreeing to be my promotor, it was a gamble, and I am glad you took it.

David, you were the first person who took time to really talk to me about my projects. Via mail or in person, you were always there. You were also the one who introduced me to the field of mathematical psychology, which further sharpened the whole thesis. I will always be grateful to you for this. Thousands of questions, quirky arguments, you could always put things into perspective. Your support and encouragement helped me a lot, thank you.

Olympia, your proper way of doing science will forever be an example of scientific rigor and good academic practice for me. Your insight on scientific methods and analysis was truly valuable. You made my writing concise and to-the-point, thank you.

Vitoria, you adopted me as your PhD student, and it was truly special to me. You gave me a combination of critical feedback and academic freedom that triggered my thoughts and pushed me to be productive. I am grateful for the multiple opportunities you offered me such as writing a paper with you and going to a CNS conference. Those were the things that I needed to experience before finishing my PhD. I also appreciated the DCC lecture we organized together; it was a great experience, both organization-wise and scientific-wise. You are a great mentor, inspiring and stimulating, patient but never pushy. I feel that I will never be able to thank you enough.

Members of the Language Function and Dysfunction group - Arushi, Johanna, Natascha, Ileana, Nikki, Adrian, Joao, Yang – thank you for the interesting discussions and the talks related and not-so-much related to science topics. Our group dinners, joint cooking evenings (go team bigos, go!) with board games and Vitoria's contraband tropical fruits truly made my time with you special.

Donders Wonders blog team – Marisha, Mahur, Roselyne, Harriette, Jeroen, Monica, Annelies, Marpessa – I really enjoyed our team work on the blogs! I think we did a great job of contributing to the blog by making it more lively, adding some controversial and difficult topics to the platform. It was a joy to see the blog go through the different stages of preparatory work, and people working on it: buddying it, editing it, translating it. It created such a feeling of comradery and accomplishment that I really needed at that time. Thank you!

Fellows from Language in Interaction Consortium – Chara, Marpessa, Marvin, Jana, Ileana, Luis Miguel, Samira, Merel, Linda, Alessandro, Izabella, Stephanie, Shruti, Tobias and many others who joined the consortium through the years – thank you for the company during multiple LiL Heidagen and Writing retreats, it was a pleasure to get to know you and your work better.

To the social communities of DCCN/DCC who organized the sailing retreats – thank you! It was one of the most wonderful experiences of my life. It is quite a feeling, to learn how to fly over the waves over the course of just two days. Thank you Elisa, for attending that retreat too, we became friends and collaborators since then. I hope the political project of ours will see the light soon!

The Donders Devils running team, you are all amazing! Experiencing Night Run, BruggenLoop, and of course 7Hills was a challenge but every swollen knee accident was worth it. Special thanks to Tildie, who hosted the 7Hills team in her house before and after the run, I thought that was very gezellig!

Alessia, Antonella, Leslie, thank you for our trips to Naples and to the Wadden Sea for the mudwalk, both picturesque and fascinating as they were! These are really great memories that I am keeping with great affection.

Natalia, your entrepreneurial mindset taught me that every piece of knowledge, no matter how small it is, deserves to be disseminated. Thank you! You were also an example of stamina and persistence with regards to the publishing process.

Stan, before leaving academia, you were my officemate, and the way we were working on that small MEG project really showed me what a pleasure it is to work on something together and to see the results of this work. You made me feel seen and after you were gone, this memory triggered me to reach out to people for collaboration and supervision. I want to thank you for that.

Xiaochen, thank you for our jogging times and for developing my palate with pig feet, I think our joint activities enriched my life.

Jessica, thanks for showing me the power of making people feel seen with words.

Arushi, thanks for our walks and talks in pre-pandemic and pandemic times, I am happy I was able to get to know you better and share some stories with you!

Johanna, Fenny, Katrin - it is always a pleasure to spend time with you while walking along the Waal, or through the city or just over the cup of coffee talking about everything and anything. You really made my days brighter.

Pessa, I miss you much. I hope I'll be able to see you one day. You were kind to me and supported me in many ways that you didn't even know.

Ronny and Mila, my human-and-dog housemates, you made this pandemic more bearable and filled the house with joy and hair, lots of dog hair.

Special shout-out thanks to the community of Nijmegen: multiple gyms, ER room, police station and immigration services. You worked as a Swiss clock and were there when I needed you. This is something not to take for granted. To multiple editions of InScience festival and Kunstnacht, that was really something special to be grateful for. I would also want to thank the admins and secretaries at the DCC and DCCN: Tildie, Sadra, Saskia, Vanessa and Mike from the TG team at DCCN. Another dank word goes to Paul Gaalman for support with fMRI data collection.

I want to thank the company Yoast, who showed me the life outside of academia, and made me think about the implementation of Scrum and Agile principles in academic environment.

A special thanks to my mother, whose cheerful chatting and never ending optimism left me no choice but to start developing the inexhaustible optimism myself.

Duru, you inspired me to finish this work and boosted the feeling that life is too short to spend it all on a thesis. For that I owe you a special thanks!

Curriculum Vitae

Lara Todorova was born in Moscow on 11 January 1988. She obtained her Bachelor's degree in Theoretical and Applied Linguistics at Russian State University for the Humanities in 2012. On her way to completing the Bachelor studies she took a gap year to participate in Erasmus program, which took her to Bologna, Italy. There she was introduced to the Foundations of Cognitive Psychology, a course taught in Italian for local students. Inspired by this experience, she moved to Rovereto, Italy, to enroll in the Cognitive Science Research Master at the University of Trento. In 2013, she moved to Nijmegen, the Netherlands, to work on her internship project at the Donders Institute for Brain, Cognition and Behaviour. Within Prediction and attention lab, Lara studied the nature and dynamics of visual mental imagery. She completed her internship and graduated from the University of Trento in 2014. In 2014, Lara started her PhD project at the Donders Center for Cognition, within the Language in Interaction Consortium. She studied the influence of language on human decision making using a combination of formal models and empirical research. This thesis is the result of her PhD investigation.

Publications

Todorova, L., & Neville, D. A. (2020). Associative and Identity Words Promote the Speed of Visual Categorization: A Hierarchical Drift Diffusion Account. *Frontiers in Psychology*, *11*, 955. https://doi.org/10.3389/fpsyg.2020.00955

Todorova, L., Neville, D. A., & Piai, V. (2020). Lexical-semantic and executive deficits revealed by computational modelling: A drift diffusion model perspective. *Neuropsychologia*, *146*, 107560. https://doi.org/10.1016/j.neuropsychologia.2020.107560

Todorova, L., Neville D.A. & Colizoli, O. When task matters: the drift diffusion perspective on language-guided face gender categorization (in prep.)

Todorova, L., Neville D.A. & Piai V. Post-perceptual effects of words on ambiguous visual perception: an MEG study (in prep.)

Bielczyk, N. Z., Piskała, K., Płomecka, M., Radziński, P., **Todorova, L.,** & Foryś, U. (2019). Time-delay model of perceptual decision making in cortical networks. *PLOS ONE*, *14*(2), e0211885. https://doi.org/10.1371/journal.pone.0211885

Mostert, P., Albers, A. M., Brinkman, L., **Todorova, L**., Kok, P., & de Lange, F. P. (2018). Eye movement-related confounds in neural decoding of visual working memory representations. *ENeuro*, *5*(4), 401–418. https://doi.org/10.1523/ENEURO.0401-17.2018

Research Data Management

This research followed the applicable laws and ethical guidelines. Research Data Management was conducted according to the FAIR principles.

Ethics

This thesis is based on the results of human studies, which were conducted in accordance with the principles of the Declaration of Helsinki. The Ethical Committee of the faculty of Social Sciences (ECSS) has given a positive advice to conduct these studies to the Dean of the Faculty, who formally approved the conduct of these studies (ECG2012-0910-058). This research is funded by Gravitation Grant 024.001.006 of the Language in Interaction Consortium from Netherlands Organization for Scientific Research.

Findable Accessible

The table below details where the data and research documentation for each chapter can be found on the Donders Repository (DR) and on the Open Science Framework (OSF) repository. The data archived as a Data Sharing Collection remain available for at least 10 years after termination of the studies.

Chapter	DAC/OSF	RDC/OSF	DSC/OSF	DSC License
2	**DAC:** di.dcc.DAC_2016.00265_775	**OSF:** DOI 10.17605/OSF.IO/J53ZK	**OSF:** DOI 10.17605/OSF.IO/J53ZK	CC-BY-4.0
3	**DAC:** di.dcc.DAC_2017.00129_610	**RDC:** di.dcc.RDC_2017.00129_946	**RDC:** -	RU-DI-HD-1.0
4	**DAC:** di.dcc.DAC_2017.00108_781	**RDC:** di.dcc.RDC_2017.00108_825	**RDC:** -	RU-DI-HD-1.0
5	**OSF:** DOI 10.17605/OSF.IO/9JQXN	**OSF:** DOI 10.17605/OSF.IO/9JQXN	**OSF:** DOI 10.17605/OSF.IO/9JQXN	CC-BY-4.0

DAC = Data Acquisition Collection, RDC = Research Documentation Collection, DSC = Data Sharing Collection, OSC = Open Science Framework.

For chapter 2, research data have been stored on the networkdrive (U837130 (\\CNAS.RU.NL)), for chapters 3, 4 research data have been stored on the projectdrives (3018025.01, 3018025.02). These data were accessible to all members involved in the project. After finalization of the project 2 data were removed from the networkdrive. The data in .csv format and analyses scripts are made public via Open Science Framework.

The articles of chapters 3, 4 are still in preparation for publication. The data will be made publicly available via Data Sharing Collection, once the article has been published. It will then be shared under the RU-DI-HD-1.0 license as it contains potentially identifiable data. Informed consent for chapters 2, 3, 4 was obtained on paper following the Centre procedure. The forms are archived in the central archive of the Centre for 10 years after termination of the studies.

For chapter 5, existing research data have been reanalyzed. The data have been collected in the University of California, Berkeley. The study protocol was approved by the Committee for Protection of Human Subjects, following the declaration of Helsinki. The research data in .csv format and analyses scripts are made public via Open Science Framework.

Interoperable, Reusable

The raw data are stored in the DAC in their original form. For RDC/OSF long-lived file formats (e.g. .sav, .csv, .tif,) have been used ensuring that data remains usable in the future. For Chapter 2 and 5, a description of the experimental setup can be found in published articles. The used software including version numbers is specified. For Chapters 3 and 4, a description of the experimental setup can be found in the DCCN project proposal database (3018025.01, 3018025.02). Analyzing scripts can be provided upon request (lara.todorova@gmail.com). Upon publication of the work in Chapters 3 and 4, the analysis scripts will be made publically available via the Data Sharing Collection in the Donders Repository.

Privacy

The privacy of the participants in this thesis has been warranted using individual subject codes. The key is stored in the investigator site file (ISF) and was only accessible to members of the project who needed access to it because of their role within the project. The key was stored separately from the research data and will remain in the ISF. Data in chapters 2 and 5 are not identifiable and can be shared without restrictions. Data of chapter 3 and 4 are can be shared under the restricted license RU-DI-HD-1.0, which provides extra statements for the protection of the identity of the participants.

Donders Graduate School for Cognitive Neuroscience

For a successful research Institute, it is vital to train the next generation of young scientists. To achieve this goal, the Donders Institute for Brain, Cognition and Behaviour established the Donders Graduate School for Cognitive Neuroscience (DGCN), which was officially recognised as a national graduate school in 2009. The Graduate School covers training at both Master's and PhD level and provides an excellent educational context fully aligned with the research programme of the Donders Institute.

The school successfully attracts highly talented national and international students in biology, physics, psycholinguistics, psychology, behavioral science, medicine and related disciplines. Selective admission and assessment centers guarantee the enrolment of the best and most motivated students.

The DGCN tracks the career of PhD graduates carefully. More than 50% of PhD alumni show a continuation in academia with postdoc positions at top institutes worldwide, e.g. Stanford University, University of Oxford, University of Cambridge, UCL London, MPI Leipzig, Hanyang University in South Korea, NTNU Norway, University of Illinois, North Western University, Northeastern University in Boston, ETH Zürich, University of Vienna etc.. Positions outside academia spread among the following sectors: specialists in a medical environment, mainly in genetics, geriatrics, psychiatry and neurology. Specialists in a psychological environment, e.g. as specialist in neuropsychology, psychological diagnostics or therapy. Positions in higher education as coordinators or lecturers. A smaller percentage enters business as research consultants, analysts or head of research and development. Fewer graduates stay in a research environment as lab coordinators, technical support or policy advisors. Upcoming possibilities are positions in the IT sector and management position in pharmaceutical industry. In general, the PhDs graduates almost invariably continue with high-quality positions that play an important role in our knowledge economy.

For more information on the DGCN as well as past and upcoming defenses please visit: **http://www.ru.nl/donders/graduate-school/phd/**