The Fictive Brain: Neurocognitive Correlates of Engagement in Literature

Article in Review of General Psychology · April 2017
DOI: 10.1037/gpr0000106

2 authors, including:

Arthur M. Jacobs
Freie Universität Berlin
345 PUBLICATIONS 12,707 CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:

- Neurocognitive Poetics View project
- Neural correlates of figurative language comprehension View project
The fictive brain: neurocognitive correlates of engagement in literature

Arthur M. Jacobs¹²³ & Roel M. Willems⁴⁵⁶

1 Experimental and Neurocognitive Psychology, Freie Universität Berlin, Germany
2 Center for Cognitive Neuroscience (CCNB), Freie Universität Berlin, Germany
3 Dahlem Institute for Neuroimaging of Emotion (D.I.N.E.), Freie Universität Berlin, Germany
4 Centre for Language Studies, Radboud University, Nijmegen, The Netherlands
5 Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands
6 Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

Correspondence: Arthur M. Jacobs
Department of Experimental and Neurocognitive Psychology, Freie Universität Berlin,
Habelschwerdter Allee 45, D-14195 Berlin, Germany.
Email: ajacobs@zedat.fu-berlin.de
Abstract
Fiction is vital to our being. Many people enjoy engaging with fiction every day. Here we focus on literary reading as one instance of fiction consumption from a cognitive neuroscience perspective. The brain processes which play a role in the mental construction of fiction worlds and the related engagement with fictional characters, remain largely unknown. We discuss the Neurocognitive Poetics Model (Jacobs, 2015a) of literary reading specifying the likely neuronal correlates of several key processes in literary reading, namely inference and situation model building, immersion, mental simulation and imagery, figurative language and style, and the issue of distinguishing fact from fiction. An overview of recent work on these key processes is followed by a discussion of methodological challenges in studying the brain bases of fiction processing.
Introduction

Fiction does not take us outside the range of human nature into something else — "convention," or "culture," or "literary tradition." Ultimately, it's all human nature. Carroll (2012, p. 298).

People read a lot – many of us everyday – and a lot of it is fiction. Even before humans can read, through tales told or stories read aloud by others, fiction changes the way we think and feel about the world, ourselves and others. Thus, fiction is a major source shaping our brain processes (Oatley, 2016; Schrott & Jacobs, 2011; Willems and Jacobs, 2016).

However, there is no too sharp a line between fact and fiction (cf. Oatley, 1999) for human beings equipped with what Darwin considered the single most important factor in the evolution of the modern human mind: Language (cf. Carroll, 2011). Listening to or reading language and fiction can have important consequences for human experience or behavior from the earliest moments of life. Lullabies as sung fiction or bed-stories shape our imagination, motivations, emotions and expectations from the early days on. We are deeply fictive animals (Oatley, 2016) and it is thus not astonishing that proponents of literary Darwinism (e.g., Carroll, 2011) posit that the disposition for producing and consuming literature has a central function the very disposition has been “designed” for by natural selection: Literature (re-)creates emotionally charged images of our experience in the world and by means of such images we orient ourselves to the world, organize our values and motives, and thus regulate our behavior. The fiction skill or fictionality1 of humans has also been identified as the central aspect of capitalist macrodynamics giving economic decisions the necessary hold in times of high uncertainty (Beckert, 2011).

Although all of this has been thought and written about for a long time in many scientific disciplines, the questions how exactly fiction is constructed in our brains and what distinguishes it from processing/re-constructing of facts is still an issue where research is basically fishing in the dark (but see Altmann et al., 2012; 2014). Recent affective and social neuroscience studies have begun to shed some light on issues related to our main topic, though. Here we review and discuss the most relevant of those.

The paper is structured as follows: We start with discussing aspects of the ontogeny of encounters with fiction and then move to literary reading in adulthood, viewed from the perspective of the Neurocognitive Poetics Model (Jacobs, 2011; 2015a,b). In section three we look at basic processes of engagement in literature, such as immersion or aesthetic appreciation. A special section is dedicated to what is perhaps the most challenging kind of fiction: poetry. Before concluding, three further sections deal with issues of i) how literary reading can change brain processes, ii) methodological challenges in more natural and ecologically valid studies of fiction, and iii) individual differences.

Reading acquisition and the ontogeny of literary response and experience

How do human beings come to like fiction? How do they acquire something like a lyrical sense? Cognitive Neuroscience so far has not even begun to shed light on the neural bases of the

---

1 “Fictionality” in economic action is the inhabitation in the mind of an imagined future state of the world and the beliefs in causal mechanisms leading to this future state.
development of literary experiences (Jacobs, 2015c). Yet, studies investigating the neural underpinnings of written language processing in children and adolescents are informative for the present purposes. In general, these studies focus on simple word recognition tasks, but a few also used longer text segments and figurative language processing.

Regarding single word recognition, the neurodevelopmental hypothesis states that beginning readers engage a widely distributed bilateral dorsal (temporo-parietal) pathway in conjunction with articulatory recoding processes in the inferior frontal regions, while the ventral (occipito-temporal) pathway is thought to be established with increasing reading experience and left-hemispheric lateralization (Liebig et al., 2017). The dorsal pathway, hypothesized to support phonology-based reading processes (i.e., grapheme–phoneme conversion, phonological assembly), includes the posterior superior temporal gyrus and supramarginal and angular gyri of the inferior parietal lobule. The ventral pathway includes lateral extrastriate, fusiform, and inferior temporal regions hosting the putative visual word form area, and is linked to memory-based visual-orthographic word recognition. Both pathways are complemented by a left inferior frontal circuit around Broca’s area that includes inferior frontal and precentral gyri thought to play a role in speech-gestural articulatory recoding of print (cf. Martin et al., 2015).

A well-developing functioning of this basic left-hemispheric (LH) reading system is a prerequisite for pleasurable and reflective experiences with literature, while both structural and functional anomalies are associated with poor reading or developmental dyslexia, both conditions which will not favor the life-long pleasures of ludic literary reading, i.e. reading for pleasure (Jacobs, 2015c, Nell, 1988). Given that children have experiences with micro-poetry like one-word poetry or nursery rhymes from early on (Jacobs & Kinder, 2015, Jacobs et al., 2015), knowledge about these basic reading circuits is useful for future studies investigating the structural and functional neural development of figurative language processing and ludic reading. While neurocognitive studies on the ontogeny of literary response are still extremely rare, there isn’t much behavioral or neurocognitive research on the co-development of language and emotion either (Sylvester et al., 2016). However, learning more about the acquisition of affective semantics, for example, is necessary for a better understanding of the development of ludic reading (Jacobs, 2015c).

In summary, the neurocognitive bases of the processes underlying the development of ludic reading and the story-like nature of the human mind are still very much in the shadow. It is highly likely, though, that full-blown ludic reading experiences require a well-developed LH reading system that automatically and efficiently decodes the written input, thus lending time and (neural) space for more complex processes of inference, interpretation, and figurative meaning construction to unfold. A well-developed vocabulary and grammar acquired through daily verbal communication and reading episodes, ideally with increasingly rich and complex verbal materials, helps, but learning to read never stops: Achieving new levels of deep reading, e.g. deliberating and reflecting text passages through combinatorial semantic, syntactic and pragmatic unification in order to enhance comprehension and enjoyment is a life-long learning process (Wolf, 2007). Its neuronal basis lies in the fact that each millisecond gained by an efficient LH reading system allows the brain to learn to better integrate (or separate) pieces of inferential, metaphoric, analogue or affective background information and world knowledge, thus producing an endless row of ever more complete thoughts and rich feelings (Schrott & Jacobs, 2011, Wolf, 2007).
Literary reading in adulthood: A model for neurocognitive poetics

The above mentioned Neurocognitive Poetics Model unifies a set of hypotheses inspired by rhetoric, aesthetics, poetics, linguistics or neuroscience within a comprehensive theoretical framework. It allows predictions concerning factors facilitating and inhibiting affective and aesthetic processes at the neuronal, subjective-experiential, and behavioral levels that should generate further research coming to grips with the complex phenomenon of processing fiction. Given both the complexity of literary reading and the paucity of empirical data from neurocognitive studies using more natural and ecologically valid reading materials (cf. Burke, 2015; Jacobs, 2015b; Willems, 2015), the Neurocognitive Poetics Model is still „work in progress“ containing underspecified parts as well as lacunae (Jacobs, 2016). However, the available empirical evidence supports the central hypotheses of the model, such as the background-foreground, the fiction feeling, or the Panksepp-Jakobson hypotheses. The background-foreground hypothesis refers to the central claim of the model that any text offers a mixture of background elements (e.g., familiar words, themes, scenes) and foreground elements (e.g., defamiliarizing stylistic devices) which activate separate routes (immersion vs. aesthetic appreciation) characterized by differing neurocognitive processes (i.e., implicit vs. explicit processing) and reading behavior (i.e., fluent vs. dysfluent reading). The fiction feeling hypothesis states that narratives with emotional contents invite readers more to be empathic with the protagonists and immerse in the text world (e.g., by engaging the affective empathy network of the brain), than do stories with neutral contents (cf. Hogan, 2010, 2014). The Panksepp-Jakobson hypothesis submits that since evolution had no time to invent a proper neuronal system for art reception, even less so for literary reading, the affective and aesthetic processes we experience when reading (cf. Jakobson’s, 1960, “poetic function”) must be linked to the ancient emotion circuits we share with all mammals, as perhaps best described by Panksepp (1998; for review see Jacobs, 2015b,c; Jacobs, 2017).

Next, we specify the likely neuronal correlates of fiction processing in more detail than in previous versions of the model. The assumptions concerning these correlates – sketched graphically in Figure 1 – are supported by an increasing number of neurocognitive studies on fiction processing that were not all published at the time, such as Altmann et al. (2012, 2014), Hsu et al. (2014; 2015b), O’Sullivan et al. (2015) or Nijhof and Willems (2015). They should nevertheless be seen as heuristic given the relative scarcity of such studies and the necessity of replication and cross-validation experiments. They can guide future region of interest, connectivity or multivariate pattern analysis studies of fiction processing (e.g., Hsu et al., 2015a) and thus help to constrain and refine neurocognitive models. In summary, the Neurocognitive Poetics Model specifies a wealth of factors and processes that may cause beneficial effects of engagement in literature organized around the two usually divergent core processes of immersion and aesthetic evaluation/appreciation (cf. also Nicklas & Jacobs, 2017; Willems & Jacobs, 2016).

[insert Figure 1 around here]
Basic processes of engagement in literature
In the following sections we discuss in more detail the likely main neural correlates of figurative and poetic language processing sketched in Figure 1 (note that this list is not meant to be exhaustive).

a. Inference and situation model building
Making meaning of a literary text or poem requires more than comprehension of words and sentences, in particular the mental (re-)construction of the situations described by a text — situation models — hypothesized to arise through the integration of a reader’s knowledge of the world with information explicitly presented in a text (Bower & Morrow, 1996; Kintsch & van Dijk, 1978; van den Broek, 2010; Zwaan, 2015). Important ‘cognitive’ subprocesses are inferences for bridging successive events/situations, the use of background knowledge and discourse context, and pragmatic interpretations. Crucial ‘affective’ subprocesses are personal experience/resonance and knowledge about atmospheres and moods conveyed, e.g., by a poetic text, and so-called mood empathy (Aryani et al., 2016; Chow et al., 2015; Gittel et al., 2016; Hogan, 2010, 2014; Jacobs et al., 2016; Lüdtke et al., 2014; Oatley, 1999).

A special role in this process is played by the left inferior frontal gyrus (LIFG; Chow et al., 2014), the dorso-medial prefrontal cortex and the (pre)/cuneus, all areas appearing sensitive to how well a coherent structure can be built from a piece of text (Friese et al., 2008; Kurby & Zacks, 2015; Martin-Loeches et al., 2008). A related but different interpretation for the role of posterior midline structures during narrative comprehension is that the (pre)/cuneus are areas with large temporal receptive windows and hence the capacity to integrate information over extended periods of time (Lerner et al., 2011). Posterior midline structures (including posterior cingulate cortex, cuneus and precuneus) also play a special role in the coding of event boundaries, i.e., points at which a narrative proceeds from one event to the other (Zacks et al., 2001; Zacks et al., 2010). Similarly, activation in the precuneus and posterior cingulate cortex correlated with a priori defined moments in a novella signaling narrative shifts (Whitney et al., 2009). Interestingly, recent work suggests that these areas may be implicated in the recognition of structure across time more generally, so not restricted to language stimuli (Tobia et al., 2012).

Both (pre)/cuneus and anterior medial prefrontal cortex also consistently appear among the areas that are more strongly activated or connected to other regions during rest as compared to during engagement in other cognitive tasks. In so-called resting-state connectivity, these areas are considered major hubs of the ‘default mode network’/DMN (Raichle et al., 2001). However, DMN might be a misnomer for areas that are co-activated during rest: Indeed, cuneus / precuneus have been speculated to be involved in conceptual processing (Binder, 2012), or more generally in internally generated cognition, sometimes labeled stimulus-independent thought, internally focused cognition, or ‘mind wandering’ relating the DMN to the construction of mental models or simulations that are adaptive and facilitate future behavior (Buckner & Carroll, 2007; Mason et al., 2007). A recent review suggests that mind wandering can occur either with or without intention and that intentional and unintentional mind-wandering are dissociable; the intentional type likely involving activation of executive-control regions (cf. Seli et al., 2016).
It may seem that being in a resting state compared to engaging with fiction are activities which overlap only minimally. However, these descriptions have in common that they relate to narrative construction: internally focused cognition, mind wandering, and mental model construction are all forms of meaning making, or of story-telling (albeit in a non-constructed manner). From this perspective it should not come as a surprise that areas that are prominent hubs when people do not engage in a specific task (‘rest’), are also prominent areas when participants are taken in by a narrative context. The overlap between resting state (connectivity) and mentalizing tasks has been described by reference to ‘self-projection’, i.e., understanding others’ behavior by means of projecting oneself into that situation (Buckner & Carroll, 2007). These functions relate to narrative comprehension directly. ‘Rest’ means that participants are not engaged in specific cognitive operations, and as such it is contrasted to ‘active’ states of performing a task. From a different perspective ‘rest’ can be regarded a mental task (or state), but with poorly defined instructions. Perhaps, building a situation model is what humans typically do when not engaged in a particular and specific task. The ‘resting state’ thus would better be viewed as an active state of being, of narrative or situation model construction. The crucial difference between situation model building during ‘rest’ and during engagement with fiction is that in the case of fiction it is externally guided. It is the narrative which imposes a discourse model upon us, whereas during mind wandering (and related constructs; cf. Smallwood et al., 2008) the situation model is generated ‘from within’, based on our memories and experiences.

How much these processes overlap and relate to the notions of intentional and unintentional mind-wandering (Seli et al., 2016) is an intriguing issue for future research.

b. Immersion and/or aesthetic appreciation

Probably the most salient feature of fiction is its power to transport readers into a narrative / fictional world. This feeling of being in the story world is sometimes called transportation and is closely related to the psychological constructs absorption and immersion. As a look through the upcoming Handbook of Absorption (Kuijpers & Hakemulder, 2017) shows, the phenomenon that people become emotionally involved, or carried away imaginatively in fiction is multifaceted, conceptually far from being unified, and difficult to measure empirically – be it with behavioral or neurocognitive methods (for recent reviews see Jacobs, 2015b, Jacobs & Schrott, 2015, Jacobs & Lüdtke, 2017). An increasing number of behavioral studies on transportation, immersion or absorption in different media and text types – including poetry (cf. Gittel et al., 2016; Jacobs et al., 2016a; Lüdtke et al., 2016) – is complemented by a few neuroimaging studies, but the evidence for neural correlates of the psychological state of immersion is still scarce and inconclusive (for review see Schlochtermeier et al., 2015).

Still, the Neurocognitive Poetics Model makes testable predictions concerning factors facilitating and inhibiting immersive processes. Thus, Hsu et al. (2014) tested and found support for the above-mentioned fiction feeling hypothesis integrated in the Neurocognitive Poetics Model. Comparing the neural correlates of post-hoc immersion ratings for fear-inducing versus neutral passages from the Harry Potter series revealed that activity in the mid-cingulate cortex correlated more strongly with the ratings for the emotional than for the neutral passages. Descriptions of protagonists’ pain or personal distress featured in the fear-inducing passages may have recruited the core structure of pain and affective empathy the more readers immersed in the text. The predominant locus of effects in the mid-cingulate cortex suggests that the
subjective immersive experience was particularly facilitated by the motor component of affective empathy for the selected stimuli from the Harry Potter series which feature particularly vivid descriptions of the behavioral aspects of emotion.

Factors facilitating immersion according to the Neurocognitive Poetics Model are related to the processing of background information contained in fiction, in particular familiarity and fluency, heightened unforced attention, empathy, identification and fiction feelings, as well as suspense, curiosity and surprise (Jacobs & Lüdtke, 2017). Factors inhibiting immersion are related to the processing of stylistic foregrounding devices (see section d) which in the Neurocognitive Poetics Model is supported by another route than the one driving immersive processes (which can be assimilated to Barthes’, 1973, *readily mode of processing*, cf. Oatley, 2016). In general, defamiliarizing text elements that make situation model building and meaning making harder, requiring schema adaptation and broader/deeper reflection will inhibit immersive processes but increase readers’ likelihood of entering a trajectory resulting in aesthetic feelings.

c. Mental simulation and imagery

At least since Plato voiced his concerns about the evocation of images via reading (Plato, 1969, 605-606), scholars have argued about the nature of such ‘images in the head’. In cognitive science, mental imagery refers to the deliberate and explicit creation of a (visual) image without direct sensory stimulation (Farah, 1989). Visual and motor imagery are typically distinguished, and imagery and actual visual perception or motor execution are to a large extent comparable at the neural level (Jeannerod, 2006; Kosslyn, 1994).

There is however a very important difference between explicit imagery, and the more implicit generation of images in the mind when we comprehend language (Burke, 2011; Jacobs, 2016; Kuzmičová, 2014; Troscianko, 2013). Literary scholars have long recognized ‘the optical poverty of my images’ during literary reading (Iser, 1976, p. 138). Indeed if humans experienced picture-like images during reading, this would be cognitively too costly to be an effective reading strategy. Explicit imagery takes a lot of time for the cognitive system, and is generally much slower than the speed at which we read. In a direct comparison, it was shown that responses in cortical motor areas during explicit motor imagery, and during reading of action verbs, could be dissociated, suggesting different neural and cognitive computations (Willems et al., 2010). Thus, we should conceptually distinguish between mental simulation or literary reading-induced (LRI, Burke, 2011) mental imagery and explicit, deliberate mental imagery. The nature of mental simulation or imagery during literary reading remains unclear (Jacobs, 2016), but some studies have investigated its occurrence in other tasks using neurocognitive methods.

There is abundant evidence from studies at the single word or single sentence level that language comprehension leads to simulation of sensori-motor and emotional content. For instance the cortical motor system (primary and premotor cortex) is active when action-related words are read (Fischer & Zwaan, 2008; see also Taylor & Zwaan, 2008; Willems & Casasanto, 2011; Willems, Hagoort, & Casasanto, 2010; Willems, Labruna, D’Esposito, Ivry, & Casasanto, 2011). Similarly, when participants read affective words or statements that imply an emotional event, parts of the ‘emotional brain’ (e.g., amygdala, anterior insula) as well as the LIFG are activated (Altmann et al., 2012; Citron, 2012; Lai, Willems, & Hagoort, 2015; Jacobs et al., 2015; 2016b;
Kuhlmann et al., 2016; Ponz et al., 2013; Samur, Lai, Hagoort, & Willems, 2015). However, simulation during language comprehension does not occur invariantly (Papeo et al., 2008), and some current proposals suggest that sensori-motor simulation is not a necessary part of language understanding (Louwerse, 2011; Willems & Casasanto, 2011; Willems & Francken, 2012).

Speer and colleagues (2009) addressed neural correlates of mental simulation during narrative processing. They had participants read a short story – word after word in RSVP – about the day in a life of a young boy. They observed activation in cortical motor areas when participants read parts of the story related to actual movements (running, throwing something, etc.). Reading of visual motion passages was found to activate brain regions involved in coding for biological / visual motion (Deen & McCarthy, 2010; Wallentin et al., 2011). Interestingly, Kurby and Zacks (2013) found that auditory and motor simulation led to activation of modality-specific cortices (e.g., the premotor cortex) only when descriptions were embedded in a coherent story, not for single sentences outside of a story context. The latter findings suggests that within a narrative context facilitating situation model building, mental simulation may be more readily part of language comprehension as compared to more decontextualized situations like single word comprehension.

Finally, in auditory narrative processing, Nijhof and Willems (2015) discovered two types of mental simulation: motor simulation of concrete actions (as in some of the studies discussed above), and simulation of intentions, thoughts and beliefs of characters in the narrative (‘mentalizing’). Activation in the motor cortex during fragments describing concrete actions suggested motor simulation, while activation during ‘mentalizing fragments’ in parts of the mentalizing/ToM network indicated mental simulation. Note that this is not a case of reverse inference (Poldrack, 2006): The inference is based on a) mapping of mentalizing areas within the subject sample of this specific study, b) previous findings as documented in meta-analyses, and c) with reference to the content of the stories. That is, the authors knew that action events were being described at certain points in the story and related those points to neural activity in an area known to be involved in action simulation (the premotor cortex) (see Hutzler, 2014 for extended discussion of reverse inference).

Interestingly, a negative correlation between motor cortex activation (while listening to action descriptions) and medial prefrontal cortex (‘mentalizing’) activation was observed. This suggests that under natural listening circumstances some readers strongly preferred to engage in motor simulation, and did not engage in mentalizing (note that no visual simulation was tested in this experiment), while others showed the opposite pattern. The authors concluded that the results reflect personal preferences. While waiting for replication, these results serve to illustrate the potential of neuroimaging to get better insight into individual differences and preferences during literary reading, and does so importantly in a task-neutral setting: All participants listened to the stories naturally, without explicit task instruction.

It is still unclear what the nature of the images evoked during literary reading is, how exactly they influence our reading experience, and how they impact appreciation and memory for narratives (but see Hartung et al., 2016). An important avenue for future research is to add to our understanding of the impact of our propensity for simulation on our fiction experience, e.g., the
dependence of LRI simulation on personal experiences (e.g. Burke, 2011). One prediction
naturally following from this is that since personal experiences differ, readers will differ greatly in
their reliance on and preference for mental simulation during literary reading.

d. Figurative language and style
Several recent meta-analyses have looked at the neural correlates of figurative language
processing with partially mixed results (Bohrn et al., 2012a; Rapp et al., 2012; Vartanian et al.,
2012; Yang, 2014). Thus, Bohrn et al.’s (2012a) meta-analysis of 22 fMRI studies on the
processing of metaphor, idiom, and irony/sarcasm in adults revealed that areas linked to more
analytic, semantic processes (e.g., LIFG) are involved in metaphor comprehension, whereas
processing irony/sarcasm more involves mPFC activation, a key region of the mentalizing/ToM
network typically involved in story comprehension (Altmann et al., 2012; 2014; Mar, 2011). In
this meta-analysis, the following key regions were associated with figurative language processing:
LIFG (BA 45/46/47) extending to the anterior insular cortex, the right IFG (BA 45/46/47),
right STG, the left MTG (BA 21/37), the medFG (BA 10/9), the left ITG (BA 20/21) and the
left amygdala. Rapp et al.’s (2012) meta-analysis of 38 fMRI studies revealed 409 activation foci,
of which 129 (32%) were in the RH, indicating that a predominantly left lateralised network,
including left and right IFG, left, MTG/STG, or medial prefrontal, superior frontal, cerebellar,
parahippocampal, precentral, and inferior parietal regions, is important for non-literal
expressions. It should be noted that the use of figurative language processing as an umbrella term in
such meta-analyses does not mean that the authors think that idioms, proverbs, conventional
and novel metaphors or other pieces of text requiring non-literal interpretation involve identical
neurocognitive processes (see, e.g., Table 1 of Jacobs et al., 2016b, for different neuronal
structures involved in idiom and proverb processing).

Since the LIFG was activated in the majority of analyses, discussion of its multiple functional roles
for literature processing is in order (cf. Rapp et al., 2012). Indeed the LIFG may be involved in
various cognitive operations. The first is meaning integration: Its anterior–inferior part may play a
key role in integrating words into meaningful supralexical units (metaphors, phrases, similes,
sentences) so that activation reflects higher cognitive demands to integrate non-literal meanings,
as opposed to literal ones, into a context (e.g., Nagels et al., 2013). Indeed one can observe a
gradual increase in LIFG activation with increasing meaning making efforts in literal vs.
metaphoric noun-noun compound (NNC) processing (Forgács et al., 2012). In Hagoort’s (2005)
Memory, Unification, Control (MUC) model, the LIFG is responsible for a unification gradient: the
interactive and concurrent integration of various word or text aspects (e.g., phonological,
syntactic, and semantic) into a coherent complex whole or meaning gestalt (Iser, 1976; cf. Jacobs,
2015b). The psychological construct of (verbal) working memory (WM) is an integral part of this
system, as the neural requirements of the unification include keeping the lexical building blocks
activated.

The second cognitive operation of the LIFG in non-literal language processing is meaning selection
and evaluation. To comprehend figurative text, it is necessary to decide whether the meaning of its
constituent words is intended to be literal or not. Thus, comprehension of the (German) idiom
“auf Wolke sieben schweben” (to float on cloud seven) likely involves a decision to read the
phrase figuratively: otherwise, the phrase will “make no sense” (Citron et al., 2015). Research on
literal language indicates that BA 45/47 may indeed regulate the selection among multiple competing responses during sentence comprehension. For example, Turken and Dronkers (2011) argue that ‘reciprocal interactions’ between BA 47 and the left MTG play a key role in selecting correct meanings, sustaining it in WM throughout sentence processing and integration into context. The same mechanism could play a role in selection between literal and non-literal meanings (Rapp et al., 2012). A third cognitive operation of the LIFG during (non-)literal language comprehension is world knowledge integration into sentence contexts (Menenti et al., 2009; Tesink et al., 2009) and stories, i.e. the situation model building discussed in section (a) above (e.g., Chow et al., 2014). A fourth operation has to do with affective meaning integration and the fact that words and texts are emotion-inducing stimuli (for review: Citron, 2012; Jacobs et al., 2015). Using NNCs coupling nouns of opposite valence (e.g., BOMB–SEX, DEATH–LUST), thus creating bivalent words creating a decision conflict in a valence decision task (Jacobs et al., 2015, 2016b), Kuhlmann et al. (2016) correctly predicted increased LIFG activation for bivalent as opposed to monovalent NNCs (i.e., NNCs composed of two positive or negative words, such as EROTIKENGEL/erotic angel or LEPRAELEND/leper misery).

Integrating the valence of several words into an affective meaning gestalt (Lüdtke & Jacobs, 2015) may involve an even more basic operation involving deeper and older brain networks such as the limbic system (cf. Bohrn et al., 2012b, 2013). Neurocognitive results concerning the liking and beauty of verbal materials reviewed by Jacobs et al. (2016b) indeed suggest that word/text valence is a compound superfeature neuronally computed at the so-called tertiary (i.e., neocortical) level of affective processing according to Panksepp’s (1998) hierarchical theory of emotions. In contrast, discrete emotions like joy/happiness and disgust appear to be more basic and central affective responses likely being computed at the secondary level (i.e., the limbic system). The neuroimaging results from Briesemeister et al. (2015) indicate that words associated with joy produce reduced brain activity in the amygdala, i.e., at the secondary level of Panksepp’s theory, while words that have positive valence, but are not associated with the basic emotion joy/happiness activate the orbitofrontal cortex at the tertiary level of affective processing.

To wrap up, while LIFG appears to be involved in many mental operations, it plays a key role in figurative language processing including affective and cognitive meaning integration, world and context knowledge, selection, and evaluation all being essential to engagement in literature. This does not mean that a wonderfully rich, subtle, and complex phenomenon like literary reading can be reduced to the well-functioning of a single brain structure; only that LIFG activation can be used in neurocognitive studies of engagement in literature as a special ROI and an index of sensitivity to figurative meaning making and aesthetic appreciation, both being closely connected according to the Neurocognitive Poetics Model (Jacobs, 2015b), e.g., via an effort after meaning dynamic (Pelowski et al., 2016). The process of closing meaning gestalts during literary reading requires slowed down eye movements, thinking and feeling, because the multitude of meaning potentials, the author has subtly created, allows to discover or construct various new ones (Iser, 1976). The reward for this increased effort comes at the end of the aesthetic trajectory: after initial moments of familiar recognition, followed by surprise, ambiguity, and tension, the closure of meaning gestalts and tension, full of relish, results from processes of integration and synthesis, occasionally supplemented by an AHA experience (Qiu et al., 2010) or feeling of good fit, ‘rightness’,
or harmony which accompanies an aesthetic feeling motivating to continue to read (Mangan, 2008; Jacobs, 2011; Kintsch, 2012).

e. Fact versus Fiction

Although fiction can feel very real, readers in the back of their heads always realize that a fiction story is just that: a creation coming from the mind of a writer, something which is made up. In a recent on-line experiment, Hartung et al. (in revision) had participants read short stories and rate their appreciation and immersion using standardized questionnaires. The stories were presented either as being written by a young writer (fiction: ‘He writes short fictional stories that are inspired by his imagination’) or as being written by a young columnist (fact: ‘He writes about his everyday life, always inspired by a real event’). Despite a large and diverse sample (N>1800) no effects were observed of the belief of the reader in whether the text was fact or fiction on their immersion or appreciation for the stories. The authors argue that differences in reading behavior may be more driven by genre expectations (newspaper versus novel) than by fact versus fiction per se when it is manipulated within the same (or similar) genre.

An fMRI study on this topic did reveal interesting differences between short stories believed to be real or not. Altmann and colleagues (2014) had participants read short narratives and labelled the stories as either ‘real’ or ‘invented’. There was large overlap in areas activated in both readings of the stories, but also critical differences: Activations in motor areas for texts labelled FACT suggested ‘an action-based [...] reconstruction of what happened’ in the story. Reading the same texts as FICTION, i.e. on the assumption that they refer to fictional events such as those narrated in a novel, a short story or a crime story selectively engaged an activation pattern comprising the dACC, the right lateral FPC/DLPFC and left precuneus, which are part of the fronto-parietal control network (Smallwood et al., 2012) as well as the right IPL and dPCC, which are related to the default mode network. The lateral frontopolar region has been specifically associated with the simulation of past and future events when compared to the recall of reality-based episodic memories (Addis et al., 2009). This suggests a process of constructive content simulation taking place during fictional reading.

In summary, the results of Altmann et al. (2014) support the assumption that reading fiction invites for mind-wandering and thinking about what might have happened or could happen. Such simulation processes require perspective taking and relational inferences which make a coactivation of ToM and empathy related areas likely. Importantly, in this study, also a personality factor co-determined neural responses to fact vs. fiction: the score on a ‘fantasy’ scale which assesses the individual tendency to put oneself into fictional characters. A stronger ‘readiness’ of readers to transpose themselves imaginatively into the feelings and actions of fictitious characters in books, movies and plays indeed lead to a stronger coupling between FPC and mPFC activity.

The distinction between fact and fiction is obviously very relevant in real life: some things we wished were real, for others we are glad they are fiction. The power of fiction lies partially in how real it can feel, and the mixed results we described in this section suggest that fact – fiction is not always a determining factor in immersion and comprehension of narratives. An engaging style, different reading goals, and – importantly – content of narratives will determine how
important it is whether something has really happened or not for how it is perceived by the reader (cf. van Krieken et al., 2015). Genette (1991) discusses five aspects theoretically allowing to discriminate between factual and fictional texts (i.e., order, speed, frequency, mood, and voice) and concludes that if at all, mode – i.e. internal focalization or direct access to the subjectivity of characters by describing their thoughts, intentions, feelings, inner dialogues etc. – is the most likely candidate. Thus, future empirical studies of fact vs. fiction processing should carefully control and/or manipulate these different aspects, in particular mode, and additionally take personality variables into account.

Poetry in the brain
Poetry is perhaps the most challenging kind of fiction, potentially revealing new layers of meaning at each and every re-reading act (Schrott & Jacobs, 2011). There is an awakening interest in the neuroscience of poetry reception and production (e.g., Chen et al., 2016; Keidel et al., 2013; Liu et al., 2015; Obermeier et al., 2016; O’Sullivan et al., 2015; Zeman et al., 2013) which we discuss in this section. Neuronal correlates of processing poetic (vs. non-poetic) texts are the bilateral precentral and IFG, as well as the right dIPFC extending into the anterior insula, and beyond to the TP. Interestingly, the dmPFC showed deactivation during reading of poetic pieces, compared to the reading of prosaic pieces (O’Sullivan et al., 2015). Further areas apparently specifically related to poetry reception are the right posterior/mid-cingulate, parahippocampal and left STG, as well as bilateral hippocampus (Zeman et al., 2013).

In their innovative comparative neuroimaging study, O’Sullivan et al. (2015) used well-construed four-line poetic vs. prosaic pieces presented incrementally, line after line, in the scanner. Their aim was to uncover the neural bases of literary awareness, i.e., the capacity to consider, manipulate, and derive meaning in complex texts which involves a more flexible situation model building process for accommodating varying related meaning threads, sensitivity to subtle meaning differences, as well as augmented social reasoning skills (likely based on empathy and ToM). According to the authors, PCC activation is related to the extent to which a situation model has been updated, ATL activity is believed to store the narrative of a situation model, and dmPFC „forces“ attention to settle on a narrative for a particular (mental) simulation. Moreover, activation of TPJ and surrounding ventrolateral parietal areas is believed to indicate reflexive updating of situation models in line with information retrieved from memory, while left IFG is thought to maintain contextual separation between representations that are similar, such as in metaphors (see above). Texts with evolving meaning are supposed to activate vmPFC – likely reflecting the motivational significance of the developing meaning – as well as lateral anterior PFC thought to be involved in construing relationships between less directly related words/meaning threads.

To sum up, in line with the results of O’Sullivan et al. (2015) in our Figure 1 three larger networks are assumed to cooperate in the meaning making of texts:

• the DMN, especially the PCC, dmPCF and ATL nodes
• the WM network including the dIPFC and superior and posterior parietal nodes, and
• the salience network including the putamen and left dorsal caudate nucleus.
The latter’s activity was triggered by Shakespearean functional shifts (e.g., I believed you were a saint; you have unhappied me by showing a bad nature) that required individuals to reason about a familiar word, and its context, in a novel way (Keidel et al., 2013). Specifically sensitive to the poetic texts in that study was a cluster of voxels that spanned from the right dorsal caudate to dACC, and further to medial and lateral anterior PFC. A continued increase in the extent of activation in IFG and LOC while readers were reflecting on poetry may indicate that they were appraising varying meanings. As concerns literary awareness, co-activation of dIPFC, IFG, temporo-occipital regions, and ATL during reading of poetry relative to prose suggests that poetic texts require the representation of multiple meaning threads (IFG), needing more focused attention during processing (dIPFC; temporo-occipital regions) in order to generate a holistic model of meaning (ATL). The observed deactivation in dmPFC, in the context of increased activation in dIPFC and AI, suggests that processing of poetic content requires a switch away from stored representations to build meaning from a novel external source. Finally, deactivation of multiple regions that typically co-activate in the DMN during the reflection phase (i.e., during 8 seconds readers reflected upon each piece of text in the scanner) potentially points to the longer time needed to establish a stable representation of meaning for poetic pieces.

How does literary reading change brain processes: the example of mentalizing and empathizing

As we noted above, it is a long-standing hypothesis that engaging with fiction can serve as a training mode for real life (Mar & Oatley, 2008; Oatley, 2016). The main proposal is that engaging with fiction trains social cognition in readers (e.g., Bruner, 1986; Gerrig, 1999; Mar & Oatley, 2008; Oatley, 2016; Willems & Jacobs, 2016). Readers of fiction make inferences (implicit or explicit) about characters’ intentions, beliefs and more generally speaking, their mental states. By doing so they implicitly train the ability to ‘step into someone else’s shoes’, an important trait for humans as a species living in a rich and often complex social environment. The abilities that have been focused on most are empathizing and mentalizing which we regard here as separable but related constructs (Kanske et al., 2016). Recent behavioral evidence indeed suggests that engaging with fiction (such as written narratives) is positively correlated with empathizing and mentalizing skills. For instance, fiction exposure was positively correlated with performance on the ‘Reading the Mind in the Eyes’ test, a validated measure of recognition of mental states (Mar et al., 2006). Others have argued that reading of a literary narrative can also lead to a direct increase in mentalizing skills, as opposed to the hypothesis that the effect of fiction reading is one that builds up over the course of development into a relatively stable personality trait; but this direct effect has been contested recently (Kidd & Castano, 2013; cf. Panero et al., 2016; Samur et al., 2017).

Neuroimaging is a promising tool to investigate the link between mentalizing and fiction reading since one can rely on a well-established set of regions known to be activated by mentalizing and empathizing tasks. The so-called mentalizing or ToM network is functionally separable both from the empathy network (Kanske et al., 2016, p. 201) as well as from the neural network involved in the basic aspects of language comprehension, such as semantic and syntactic processing (Willems & Varley, 2010). Despite its promise, the available evidence from neuroimaging for a link between engaging with fiction and changes in neural make-up is limited, especially as concerns developmental aspects. In an innovative neurocognitive study on the
development of cognitive and affective empathy in auditory story processing, Brink et al. (2011) found that empathizing with a character not only entails understanding why the other person is happy or sad (i.e., cognitive empathy), but also the ability to experience these emotions with her or him (i.e., affective empathy). With increasing age (4 – 8 years) activation in medial OFC, left IFG, and left DLPFC increased for the affective empathy conditions suggesting that these areas play a role in age-dependent shifts in affective empathy possibly co-occurring with maturation of the above-mentioned fronto-subcortical circuits and the development of the ToM network. Thus, a facilitatory factor for later episodes of ludic reading may be the acquisition of good mentalizing abilities associated with the well-functioning of neuronal ToM networks and domain-general nodes of the DMN (Aboud et al., 2016), the development of which, in turn, is facilitated by reading fiction (Mar, 2011; Oatley, 2016).

Regarding adult readers, parts of the neural network involved in social cognition were shown to be more strongly activated during comprehension of brief excerpts of fiction related to social content in those that engaged with fiction more (Tamir et al., 2016). Willems and Hartung looked at differences in correlations of time courses between regions while participants listened to literary narratives or a reversed speech version of the same stories – which served as a low-level baseline – to investigate the influence of self-reported amount of fiction reading on these (Willems & Hartung, under review). The outcome was that several regions turned out to be connected to many more other regions during listening of the narratives in those that reported to read more as compared to those who reported to read less. Key regions showing increased connectivity in avid readers are: inferior frontal cortices bilaterally, lingual gyri bilaterally, right middle frontal gyrus, posterior part of SMG, and anterior part of the MFC. These include regions that are part of the mentalizing and language networks, and hence these data can be taken as tentative support for the hypothesis that regularly reading fiction trains the language network (unsurprisingly perhaps) as well as the mentalizing network.

Although these neurocognitive studies provide correlational evidence only and the relation between lifetime reading and ToM is also correlational (and, presumably, bidirectional) hints to direct causal effects have been found in behavioral experiments summarized in Oatley (2016): for example, better performance in objective empathy tests or subjective self-reports for participants in fiction vs. non-fiction reading groups, as we briefly outlined above. Interestingly, the behavioral study by Bal and Veltkamp (2013) suggests that a potentially causal effect of fiction on empathy may be mediated by emotional transportation into the story.

**Methodological challenges**

**Neuroimaging using continuous stimuli**

A commonly named hindrance to applying neuroimaging to the study of literary reading is that fMRI cannot be used with continuously presented stimuli. It is common ‘wisdom’ that in using fMRI, stimuli have to be presented with a considerable intertrial interval of several seconds making the technique less suitable for use when participants read or listen to longer stretches of narrative. However, modern analysis techniques make it possible to use while participants read or listen to narratives presented at a natural pace. The reason why continuous stimuli are typically avoided in fMRI has to do with the slowness of the BOLD response. If one presents
stimuli in rapid succession, BOLD curves to each stimulus start to overlap, and it’s difficult to assess which stimulus generated which response. Using an RSVP variant Yarkoni et al. (2008) had participants read short narratives one word at a time, presented for 200 ms, with an inter-word interval of 150 ms. Estimating the BOLD curves to several word characteristics, they asked, e.g., which brain regions were sensitive to differences in lexical frequency between words, variation in the latter creating the necessary variance in the estimated BOLD curve. In a similar vein the BOLD response to action and mentalizing events was modeled within a narrative presented auditorily at a normal speech rate (Nijhof & Willems, 2015). While in the auditory modality this rapid serial presentation is the natural mode of processing, the reading results by Yarkoni et al. (2008) or Speer et al. (2009) require replication with materials that are read at a more natural reading speed. Note that this is technically feasible as studies show which successfully combine eye movement measures (eye tracking) with fMRI (Choi et al., 2014; Schuster et al., 2016).

Another way of analyzing fMRI data that are acquired while participants engage in viewing or listening to continuous stimuli is to present short narratives scrambled in time at different time scales. Participants listen to the original (no scrambling), to a version in which paragraphs are scrambled (breaking continuity at that particular time scale), to a version in which sentence order is scrambled, or a version in which words are scrambled. (Lerner et al., 2011). Inter-subject correlation analysis can then be used to assess which brain regions show a similar time course across participants for the original story, comparing this to brain areas which show the same time course across participants for the scrambled versions (for other analysis techniques see Andric & Small, 2015).

**Individual differences**

Another methodological challenge for neurocognitive – or more precisely, all – studies of literary reading are individual differences. It is well known, for example, that cognitive variables such as WM span or vocabulary scores co-determine speed and accuracy of language comprehension as can do affective variables such as mood (Van Berkum et al., 2013). More generally, it is often remarked (but not empirically investigated) that individual differences in brain responses would increase once researchers start investigating language processing at the discourse level, and the above mentioned studies by Altmann et al. (2012, 2014) or Nijhof and Willems (2015) lend support to this argument. This can be considered a nuisance if one adheres strongly to a research tradition which focuses on explaining common variance within a research sample. In this tradition, effects which can be observed reliably across the sample are taken to be reliable effects which can be extrapolated to the population level. Crucially, individual differences hinder common group effects since they are not observed across the sample but at the level of the single reader. As is well known, psychology has a rich tradition of investigating individual differences, and there is no formal reason why the approaches developed in the past cannot be used in cognitive neuroscience as well.

One problem, though, are the traditionally very low sample sizes in neuroimaging (mainly driven by the high costs). The current trends of increased sample sizes and data sharing could provide an impetus for more commonly looking at individual differences, and we have indeed quoted several studies in this paper that combine fMRI data with measures of individual differences. A
related development is the increasing popularity of doing statistical analysis with the help of linear mixed or hierarchical drift diffusion models which allow for flexibly and explicitly testing individual differences (e.g., Lüdtke et al., 2014; Froehlich et al., 2016; van den Hoven et al., 2016).

**Conclusion**

Neurocognitive studies on fiction constitute a small but rapidly evolving niche in cognitive neuroscience. We have outlined several areas of active investigation and showcased examples of how neurocognitive methods and models can help in understanding how we engage with narratives. It should be clear that there are only few hard and replicated facts in this still juvenile area of research. We hope our contribution facilitates identifying promising topics for future research. Instead of summarizing the points we made in the paper we would like to end with two general statements.

First, brain imaging is a tool, not a goal in itself, and within the field of (neuro-)cognitive poetics requires complementary direct measures, e.g., of experiential processes (Dixon & Bortolussi, 2015; Jacobs, 2015c, 2016; Kuiken, 2015). Neuroimaging can help in characterizing processes involved in narrative comprehension, thus aiding to understand which psychological or social constructs neurofunctionally overlap and which do not. Neurocognitive findings like those of Altmann et al. (2014), Brink et al. (2011) or Willems and Hartung (in revision) can lead to a deeper understanding of the effects of fiction, and how we engage with it emotionally and cognitively. They complement behavioral studies like those of Kidd and Castano (2013), Bal and Veltkamp (2013) or Jacobs et al. (2016a) by casting light on the ‘on-line’ microprocesses and allow to test hypotheses difficult to test with behavioral measures alone. For example, Bal and Veltkmaks’ (2013) conclusions concerning the effects of fiction reading are based on off-line, post-hoc ratings (i.e., memories) of reading entire texts (e.g., a 2750- word story). Thus, here the construct emotional transportation refers to a remembered experience (more or less vulnerable to memory decay and distortions) concerning effects of the text as a whole, at a macroscopic level (Jacobs, 2015c). If we had complementing fMRI data that, say, indicate selective recruitment of brain networks previously associated with fiction feelings and immersion (e.g., Altmann et al., 2012; Hsu et al., 2014, 2015) in the conditions yielding higher ratings, then this would increase our confidence in the rating data. More importantly, it would also allow testing more specific hypotheses (by using psycho-physiological interaction analysis or dynamic causal modeling), e.g., to what extent the ToM network, the autobiographical memory network or other networks of interest were co-activated. This, in turn, could lead to new hypotheses allowing to refine the fuzzy construct ‘emotional transportation’ (Jacobs & Lüdtke, 2017).

The long-term goal should be understanding how fiction ‘works’ in neurocognitive terms, a goal that cannot be reached without general theoretical tools such as the Neurocognitive Poetics Model and specific computational and process models (e.g., Hofmann & Jacobs, 2014; Jacobs et al., 2016b). Second, next to researchers interested in fiction per se, the empirical study of fiction is a useful arena also for those who work in seemingly distant subdisciplines of cognitive science. Fiction is a natural habitat of (among others) mental simulation and mentalizing, integration of information in memory, language comprehension, or emotion (Willems & Jacobs, 2016). Researchers interested in these topics should consider fiction as a way of performing their
studies, greatly increasing ecological validity. In the present paper we have shown that this is possible and – so we hope – worthwhile.
References


Hartung, F., Withers, P., Hagoort, P., & Willems, R. M. (under review). When fiction is just as real as fact: No differences in reading behaviour between stories believed to be based on true or fictional events.


Willems, R M, & Hartung, F. (under review). Engaging regularly with fiction influences connectivity in cortical areas for language and mentalizing.


**Figure Caption.**

**Figure 1.** Extension of the Neurocognitive Poetics Model sketching the likely main neural correlates of subprocesses involved in implicit and explicit fiction processing, e.g. situation model building, immersion, or aesthetic appreciation. Some of these structures are included in networks, in particular: the DMN (PCC, dmPCF, ATL) nodes, the WM network (dlPFC, superior and posterior parietal nodes), and the salience network (putamen, left dorsal caudate nucleus).

Abbreviations: LH = Left hemisphere, RH = right hemisphere, vOT = ventral occipital cortex, MTG = medial temporal gyrus, IFG = inferior frontal gyrus, AG = angular gyrus, SPL = superior parietal lobulus, SMG = supramarginal gyrus, ATP = anterior temporal pole, A/PmCC = anterior/posterior/medial cingulate cortex, TPJ = temporo-parietal junction, dlPFC = dorsolateral prefrontal cortex, Ins = Insula, OFC = orbitofrontal cortex.

**Predominantly LH: implicit processing, fluent reading**
- Left panel. Ventral & dorsal reading paths: vOT, MTG, IFG, AG, SPL, SMG
- Right. Situation model building & updating /schema assimilation: ATP, PCC, Precuneus; Fiction feelings-immersion, (vicarious) fear, anger; care: mCC, Amygdala, ACC.

**LH & RH: explicit processing, aesthetic feelings, dysfluent reading**
- Left. Figurative meaning computation, situation model/schema adaptation, augmented attention: IFG, TPJ, dlPFC
- Right. Lust, play, seek, aesthetic trajectory: Ins, OFC, caudate, hypothalamus.