

Fantastic Cognition

Karin Kukkonen

Our brains build models of the world and continuously modify these models on the basis of the signals that reach our senses. So, what we actually perceive are our brain's models of the world. They are not the world itself, but, for us, they are as good as. You could say that our perceptions are fantasies that coincide with reality. (Frith, 134-135)

Put simply, the brain is – literally – a fantastic organ (fantastic: from Greek *phantastikos*, able to create mental images, from *phantazesthai*). (Friston, 1328)

Reality or dream? truth or illusion?

Which brings us to the very heart of the fantastic. (Todorov, 25)

According to recent developments in predictive processing and Bayesian cognition, our thinking is fantastic: we grasp the world through predictive, probabilistic models that we compare against the feedback from the actual environment, which really only plays a significant role if it proves our predictions to be wrong and forces us to revise the probabilistic model. Rather than our eyes and brains registering every detail of the environment and configuring it into the larger whole of the percept, perception works the other way around. We already know what we are likely to perceive on the basis of our predictive probabilistic models (or ‘fantasies’) and correct these only if discrepancies with the environment create prediction errors. The predictive processing model of cognition, with its feedback loops of virtual models and prediction errors, has in recent years been extended into many other areas of the cognitive sciences, from perception to motor control, from emotions to our sense of self (see Clark and Hohwy for comprehensive overviews). Not only perception but cognition more generally might turn out to be nothing short of fantastic.

The predictive, probabilistic models of perception and cognition which Frith and Friston gloss as “fantasies” are, as we shall see, usually not noticed in everyday life. Tzvetan Todorov’s definition of the literary genre of “the fantastic”, as a genre that makes us hesitate between “truth” and “illusion”, however, suggests that literary texts might often test boundaries and highlight the workings of such cognitive models. As I will go on to show, the fantastic in literature throws “fantastic cognition” into relief.

* * *

Karl Friston, one of the leading proponents of the predictive processing model in the cognitive sciences, makes his statement about the brain as a “fantastic organ” in a place that is perhaps unexpected: a review of a book on art. Eric Kandel’s *The Age of Insight: The Quest to Understand the Unconscious in Art, Mind and Brain, from Vienna 1900 to the Present* (Random House, 2012) begins by painting a lively picture of the intellectual scene in Vienna around the turn of the twentieth century, where psychologists exchanged views with the most progressive artists, such as Gustav Klimt, Egon Schiele, and Oskar Kokoschka. From this, Kandel moves into a more general overview of current neurological investigations into the perception of art, in particular its visual properties (as in the work of Semir Zeki) and its empathy-evoking capacities (as in the work of Chris and Uta Frith). He makes a double

claim about the interaction between visual arts and the cognitive unconscious. On the one hand, the formal features of artistic style exaggerate what we would see in the real world (think of Klimt's gold-studded portraits), and thereby guide the perceiver's attention in particular ways. On the other hand, the creative work of artists can call attention to the usually unconscious processes of perception, in particular to what Hermann von Helmholtz called "unconscious inference" (through which we match predictive models with the visual input before actually perceiving something).

In a way, Kandel's book recreates the Kaffeehaus exchanges between science and art by bringing together a rich catalogue of the work of Klimt, Schiele, and Kokoschka with a very detailed account of the latest research on the cognitive processes at play when perceiving such paintings. Friston chooses to home in on one particular aspect of this exchange: the role which expectations play in the perception of art. This is not surprising, given that he is the champion of predictive-processing approaches in psychology. Kandel's book reminds us also, however, that predictions and expectations have played a central role in art theory for a long time. The work of Ernst Gombrich, for instance, who himself has Viennese roots, brings together art perception with the psychological research of J.J. Gibson and William James in his seminal *Art and Illusion: A Study in the Psychology of Pictorial Representation* (1959). Gombrich foregrounds "the beholder's share" in the co-construction of artistic images. The paintings and statues that most successfully engage viewers are usually not those with the smoothest finish and the most precise detail, but rather those that leave the beholder space to complete details and thereby engage with the work in more cognitively profound ways. Artists provide, for example, blurry pictorial elements that lead perceivers to reconstruct more vivid details (Gombrich mentions here the late work of Rembrandt and the Impressionists) or simplified outlines of facial features that let perceivers focus on the emotional import of the expression. Because perceivers know from the context (of the man and the coat in, say, Rembrandt's *Portrait of Jan Six*) that they *should* perceive a set of gold buttons where there are blurry lines and knobs of gold and black on the canvas, they do – and, according to Gombrich, they imagine a richer play of lights reflecting off these buttons to boot.

To some extent, Gombrich reimagines the history of (mostly Western) art as a history of artists devising ever more refined tools to engage the beholder and to put her "share" in the image to ever more sophisticated uses. They do not strive toward an exact representation of reality, though. Instead, artists (often wilfully) engage in visual fraud and use short-cuts of representation that distort what perceivers would actually see in the real world, but at the same time engage our schemata of perception (or rather, our expectations) more efficiently. Not just in the real world, as work on inattention blindness and change blindness suggests (see O'Regan and Noë), but also in the perception of art, we fall prey to surprising lapses in perception. Perceivers do not notice impossible mirror angles, contradictory foreground-background compositions or inconsistent shadowing (Cavanagh). These typical mistakes in the perception of art are exploited by artists to create engaging images (Gombrich argues). At the same time, these techniques, usually devised by the artist after long experimentation with the effect in interplay with her intuition, give scientists access to the "simplified physics" which the brain uses to recognise scenes. In a *Nature* article of 2005, Patrick Cavanagh details how the typical short-cuts in perception that I listed above depend upon the "physiology of the visual brain" (301). He makes the claim that the artistic process constitutes "a type of found science" about perceptual shortcuts and the simplifications which the brain employs when recognising objects and scenes. Indeed, in *Art and Illusion*, Gombrich discusses Constable's cloud studies. He shows how the artist draws sketches and changes their set-up systematically, as he gets a grip on the visual possibilities as an artist. This practice constitutes arguably the experimental set-up of "found science". Gombrich

writes, “I think [Constable] felt that the history of science presented a story of continuous advance in which the achievements of one observer were used and extended by the next” (175).

It seems that this discussion of Gombrich’s “beholder’s share” and the “simple physics” of artistic shortcuts have taken us a long way away from predictive processing (which indeed does not play quite as central a role in Kandel’s book as Friston makes out and which Cavanagh does not mention at all). Friston, however, is right to connect in particular Gombrich’s “psychology of art” with predictive processing. The mimesis of the work of art in Gombrich is based not on its truthful representation of the real world, but on the degree to which it engages the expectations of the spectator, thereby creating the illusion of mimesis that is both more artificial and more powerful. The viewer needs to be given both the opportunity and the means to “project what is not there”, through under-defined elements on the canvas (a “screen”), and clues for what inference to draw (Gombrich 232). The viewer’s predictive models take centre stage, as artists devise depth compositions and colour constellations which are empirically incorrect but do not create immediate prediction errors. These instances can give us insights into the make-up of the predictive model, not only for the perception of art but also for the perception of the rest of the world.

Remember Chris Frith’s suggestion, quoted at the beginning of this article, that “perceptions are fantasies that coincide with reality”. For him, the brain literally “makes up” the mind in that it constitutes the physical basis of cognitive operations. The brain “makes up” the mind in the extended sense as well, because predictive processing in the brain gives rise to cognitive illusions, such as the assumption that our minds are isolated and private. It would take up too much space to go into the details of Frith’s delightful account here, but basically, he suggests that cognition generally relies on Helmholtz’s “unconscious inferences” in visual, proprioceptive, interoceptive, and other kinds of perception. We perceive, feel, intend, and think after these unconscious inferences have related the sensory stimulus to our predictive, probabilistic models. Like Gombrich’s notion of artistic mimesis, Frith’s concept of cognition depends on expectations, understood as virtual models. Predictive models for example readjust our perception of the source of sound when watching TV from the speakers of the TV set (where it actually comes from) to the mouth of the speaking actor on screen (see Hohwy, 131). As Frith puts it, “For us to act upon the world it doesn’t matter whether or not our brain’s model is true. All that matters is that the model works” (136) – and, he goes on to add, that it matches the models of others.

Artistic styles and strategies, as Cavanagh and Kandel suggest, might offer one way for getting to the bottom of these predictive models (or fantasies) that are so important for our perception. Indeed, art historians like Gombrich who carefully trace the emergence and development of these stylistic devices (or artistic short-cuts) in light of the cognitive schemata and predictive models they engage can provide the sciences with hypotheses to test in the Bayesian paradigm. As Ladislav Kesner points out, the same might be true for Erwin Panofsky’s notion of “iconography” (conventional features of images that lead to the identification of the characters and the scene depicted), for example, and Michael Baxandall’s notion of the “period eye” (historical ideas about visual perception that find their way into the presentation of reality in the image). Indeed, according to Kesner, pre-modern works of art (such as Chardin’s *Lady Taking Tea*, which Baxandall discusses in detail, 74-104) contain their own “script for action” which art historians can trace (Kesner, 10). These pictures provide viewers both with a particular artistic vision and with in-built instructions (through compositional arrangements or references to contemporary debates around perception) for how the images should be perceived. In other words, works of art not only engage the predictive models on which (in the Bayesian paradigm) our cognition is based, but they also

offer viewers little prediction errors that cue them to refine their predictive model in a particular way.

* * *

From our discussion of visual perception and artistic styles emerges a relatively well-defined research programme for science and visual art within a predictive processing framework. Can we posit a similar kind of “found science” for literature? Within the growing field of the cognitive study of literature, the historical sciences have been taken into account for the discussion of cognitive phenomena in the texts of earlier periods (see for example Crane’s *Shakespeare’s Brain*, Richardson’s *The Neuronal Sublime*, and Anderson’s *Renaissance Extended Mind*), and there are noteworthy attempts to make sense of neurological evidence through the interpretive paradigms of the humanities. In *Feeling Beauty*, G. Gabrielle Starr discusses neurological studies of the sister arts (literature, painting, and music) through Baumgarten’s notion of the aesthetic as “a blend of sensation and knowledge such that we almost feel thought itself” (xiv). Paul B. Armstrong, in *How Literature Plays with the Brain*, combines the seemingly contradictory neurological indications of our predilection for familiarity and regularity on the one hand and novelty on the other hand through the principle of the hermeneutic circle that develops out of the interplay between what is known and what needs to be explained. Both Starr and Armstrong give rather general statements about the interpretive models which they propose, rather than tracing a specific set of cognitive short-cuts through stylistic devices (as Cavanagh proposes) or unfolding a historical narrative of the development of these devices through an archive of artistic experimentation (as Gombrich does). With the predictive processing and Bayesian approaches to literature still vastly under-represented in cognitive literary studies, the value of literary stylistic and narrative features that create particularly instructive and insightful instances of what I called “fantastic cognition” in the title of this article still needs to be asserted for literature.

Fantasy is traditionally the domain of literary study, and is the subject of many literary texts. So, in the interests of developing a “found science” of fantastic cognition from a literary point of view, let us have a look at what is generally considered the foundational text of the literary fantastic, used to exemplify Todorov’s initial discussion of the term: Jacques Cazotte’s *Le diable amoureux* (The Devil in Love; 1772/1776).¹ In Cazotte’s story, young Alvare chooses to dabble in the dark arts, and he summons a creature that first takes the shape of a camel’s head, then a dog, and later the page Biondetto. Sometimes, Alvare finds it really difficult to make out who is in front of him:

Le feu de ses regards perçait à travers le voile, il était d’un pénétrant, d’une douceur inconcevables: ces yeux ne m’étaient pas inconnus. Enfin, en rassemblant les traits tels que le voile me les laissait apercevoir, je reconnus dans Fiorentina le fripon de Biondetto; mais l’élégance, l’avantage de la taille, se faisaient beaucoup

¹ Most of the works which we would intuitively classify as “fantasy”, such as Tolkien’s *Lord of the Rings* or Martin’s *Song of Ice and Fire*, are in Todorov’s classification instances of the “marvellous”, because here wizards and dragons actually exist in the fictional world. The marvellous (*le merveilleux*) is a long-standing term in literary criticism that pertains to divine interventions, supernatural passions, and larger-than-life narratives that were not uncommon, for example, in the romances of the sixteenth and seventeenth centuries. The realist eighteenth-century novel rejects it (generally), but with the rise of the Gothic towards the end of the eighteenth century (of which Cazotte’s *Diable amoureux* can be considered a forerunner), the marvellous turns into a supernatural challenge for newly won empirical epistemic certainties.

plus remarquer sous l'ajustement de femme, que sous l'habit de page. (Cazotte, 323)²

In the singer Fiorentina, he perceives the features of Biondetto, but in turn, he considers that the female form fits Biondetto much better than his outfit as a page. (Indeed, soon afterwards Biondetto is revealed to be a seductive young woman: Biondetta).

Alvare's perception seems to be akin to the perceptual illusions such as the duck-rabbit and "binocular rivalry", some of the most discussed issues in relation to predictive processing in visual perception. In the experimental paradigm of "binocular rivalry", an image of a house, say, is presented to one eye and a face to the other, and most participants' perception switches back and forth between perceiving a house and perceiving a face, rather than combining the two into a single image. Similarly, Alvare looks at Fiorentina, but perceives the eyes and the mischief of Biondetto. At the same time, the female form seems to be more fitting for Biondetto than his original gender suggests. Alvare is unsettled by the dual percept of Fiorentina the woman, or Biondetto the young man, because these two options correspond to different predictive, probabilistic models that he has and should not coincide. In the predictive coding account of binocular rivalry, the reason why we have trouble perceiving both the house and the face lies in the fact that our expectations of the world (deeply ingrained in our bodily experience) do not allow us to posit two such conflicting percepts in the same place – no matter what the actual visual stimulus is and no matter that subjects know very well that they are in a laboratory setting, where the expectations derived from the natural environment do not necessarily pertain (see Hohwy 19-23 for a more detailed discussion). Similarly, Alvare struggles to see both Fiorentina and Biondetto in the same person, even though he knows that the supernatural capacities of the creature he has summoned would make such shape-changing quite possible.

Clearly, the similarities between visual illusions (which depend on predictive models working as they should) used in scientific experiments and in artistic devices and styles across the history of visual art are more straightforward, because they both deal with visual perception. In literary texts, written language supplies a complex web of cues for reconstructing perception, proprioception, actions, thought processes, and direct speech cognitively. The linguistic cues draw on a plethora of cognitive modes and, for the most part, they add a screen of representation to the events, actions and thoughts in the narrative that needs particular attention when matching the cognitive process to the textual example. Nonetheless, we can say that literary texts work through a similar strategy of mimesis as the one that Gombrich has identified for visual art. In the linguistic mode, too, literary texts offer readers space to respond and specify their inferences (to "project", in Gombrich's words), and they also provide the necessary cues for doing so. Both verbal and visual mimesis works through the interplay between cognition and the crafted exploitation of our cognitive predictive models through the text.³

² In English (this and the following translations are mine): The fire of her gaze broke through the veil, it was of an inconceivable penetration and sweetness: these eyes were not unknown to me. Finally, combining the features which the veil let me perceive, I recognised in Fiorentina the mischief of Biondetto; but the elegance, the shapeliness of her waist we much more remarkable in the guise of a woman than when dressed as a page.

³ Indeed, in the years before cognitive approaches to literature, such an understanding of literary art was developed in great detail through hermeneutic approaches to literature. Critics speak of "Unbestimmtheitsstellen" (spots of indeterminacy; Roman Ingarden) or "Leerstellen" (empty places, gaps; Wolfgang Iser) which the text leaves for readers to contribute to the meaning-making process. The artistic reconfiguration of reality in the literary text leads readers to revisit their predictive models and to refine their inferences in the "hermeneutic circle" (Hans-Georg Gadamer) or in the engagement with "mimesis II" (Paul Ricœur). This is a wide and complex field in the history of literary criticism, which I cannot discuss in any detail here, but these brief references hopefully go to show that a long tradition of pre-cognitive literary theory

Literary texts like Cazotte's differ from pictures not only in the verbal mode of representation but also in their explicitly narrative construction.⁴ Elsewhere (see Kukkonen), I have discussed narratives as containing a probability design, a feedback loop between the probabilities of the fictional world and the events of the plot, which create prediction errors and force readers to revise their predictive model of the fictional world. Narrative, in other words, shapes our Bayesian inferences. What does this have to do with binocular rivalry and duck-rabbits? It turns out that there is an interesting connection between work on visual illusion and narrative. Jakob Hohwy notes that when we are presented with pairs of ambiguous visual stimuli (such as two duck-rabbits), a change of belief in the situation can determine how we perceive them. A weakly narrative contextualisation, such as "the duck is about to eat the rabbit" (Hohwy 131) creates what Hohwy calls "cognitive penetrability" from higher-level propositional expectation to cognitive percept. The predictions of our conscious knowledge actually have an effect on our perception. Note that in the usual cases of visual illusion, there is no such "penetrability": we know that this is a duck-rabbit, but we only ever perceive a duck or a rabbit, because other – preconscious – predictive models apply. Hohwy cites only one study in support of this claim (Jensen and Mathewson 2011),⁵ but if he is right, this finding offers many exciting avenues for cognitive literary study and the cognitive sciences, because narrative then directly affects the interplay between conscious predictive models and Helmholtz' 'unconscious inference', and thereby shapes how we employ predictive processing. In what follows, I will outline some possible case studies through further discussion of Cazotte and the fantastic.

* * *

The ambiguity of Biondetto/Fiorentina is indicative of the epistemic destabilisation that runs through Cazotte's entire novella and that led Tzvetan Todorov, in *The Fantastic*, to posit *Le diable amoureux* as the paradigmatic example of the genre of the fantastic. According to Todorov, the key sign of the fantastic is the moment of hesitation over whether the events in the narrative can be explained through supernatural forces or through the machinations of a trickster. If the events are supernatural within the fictional world, the text tilts into the marvellous (*le merveilleux*). If they can be realistically explained, we get the uncanny (*l'étrange*). If the text continues to hesitate until its very end, as is the case with Cazotte's novella, it is an example of the truly fantastic (*le fantastique*). Todorov has defined the fantastic as "that hesitation experienced by a person who knows only the laws of nature, confronting an apparently supernatural event" (25). In other words, the fantastic is a literary genre which creates tension between competing predictive models.

In Cazotte's *Diable amoureux*, readers are invited to hesitate between two general predictive models that also describe the hesitation of the fantastic: "Alvare is seduced by the devil" and "Alvare is duped by the charlatan Soberano"/"All the events are hallucinated by Alvare". While the suspicion that all is staged by Soberano can be dismissed relatively quickly, the juxtaposition between the devil's seduction (which would make the narrative marvellous) and Alvare's dream or hallucination (which would make the narrative uncanny) remains. Each of these predictive models privileges different "unconscious inferences" for readers. If the devil actually takes the shape of Biondetta, then readers need to understand

suggests that something akin to the predictive processing account of predictive, probabilistic models that work in interplay with (designed) prediction errors can also be considered when discussing the reading of the literary work of art. But this is a research programme which will need a separate treatment.

⁴ Of course, images and in particular image sequences as in comics can be narrative, too.

⁵ Also, other stimuli used in earlier studies of binocular rivalry, such as grates with different orientations (see Hohwy 21; Fig. 3), can perhaps not as easily be turned into narratives as the duck and the rabbit.

everything that she says as quite likely being part of a strategy to seduce Alvare. Her confession that she is a sylph turned human, firstly to teach the magicians around Soberano a lesson and, secondly to be fit for Alvare's love, will have to be read as a ruse. The assassination attempt on Biondetta then is also probably set up by her herself. If, on the other hand, Alvare dreams most of this, then we need to start questioning the reliability of his narrative. Maybe he just imagines the similarities between Fiorentina and Biondetto? Does he spin his own narrative around the young woman Biondetta, perhaps in order to mask the fact that he cannot commit to her? Does he wish to become a charmer of the devil with such petulance that he simply imagines it rather than practise actual witchcraft? After all, he admits himself that "jamais rendez-vous galant ne fut attendu avec tant d'impatience"⁶ (318) as his meeting with Soberano to conjure the devil.

The two general predictive models (or "hyper-priors" in the parlance of predictive processing) reconfigure the textual evidence in conflicting ways. These predictive, probabilistic models cue readers to pick up on different elements of the text in the hope that they will confirm one hypothesis and help to disambiguate the situation. As Friston puts it, "the *raison d'être* for inference is to disambiguate among plausible and competing hypotheses" (1329). The events in the narrative, the statements of the narrators (and their slips of the tongue), serve readers as evidence to decide which predictive model to apply to their reading of the text. Such inferences are usually not part of the conscious experience in reading, mostly because there is typically only one general predictive model for the fictional world. In the feedback loop of what I call the "probability design" of the literary narrative (see Kukkonen), the plot events (and the new information about the fictional world which they carry) lead to a revision of the predictive, probabilistic model of the fictional world. New observations hence usher in a modification of the existing model that can be surprising and unusual. The fantastic (in Todorov's sense), however, systematically brings the "unconscious inferences" in literary reading to the fore because it forces readers to hesitate between contradictory models, and thereby makes the inferencing process in many instances more conscious.

Of course, not all the inferences in *Diable amoureux* are foregrounded explicitly through Alvare's narration. Consider the following example: Fiorentina appears at the dinner where Alvare is entertaining his cabalist friends, right after he mentions to Biondetto that Fiorentina had promised to attend:

"Biondetto, dis-je au page, la Signora Fiorentina m'a promis de me donner un instant, voyez si elle ne serait point arrivée." Biondetto sort de l'appartement.

Mes hôtes n'avaient point encore eu le temps de s'étonner de la bizarrerie du message, qu'une porte du salon s'ouvre, et Fiorentina entre, tenant sa harpe, elle était dans un déshabillé étoffé et modeste, un chapeau de voyage et un crêpe très clair sur les yeux, elle pose sa harpe à côté d'elle, salue avec aisance, avec grâce: "Seigneur Don Alvare, dit-elle, je n'étais pas prévenue que vous eussiez compagnie, je ne me serais point présentée vêtue comme je suis, ces messieurs voudront bien excuser une voyageuse." (Cazotte 322)⁷

⁶ In English: "Never was an amorous rendez-vous expected with such impatience".

⁷ In English: 'Biondetto, I say to the page, Signora Fiorentina has promised me a moment with her, see if she has not arrived.' Biondetto leaves the apartment. My guests had not had the time to be surprised by the strangeness of the message when a door opens and Fiorentina enters, holding her harp. She was in a sweeping and modest dishabille, with a travelling hat and a sheer veil in front of her eyes. She puts her harp down beside her, gives a poised and graceful greeting: 'Signor Don Alvare, says she, I was not warned that you would have company. or I would not have presented myself dressed like this. Messieurs, please excuse the appearance of a traveller.'

Fiorentina appears with the same promptness with which Soberano's pipe had been refilled earlier, and it seems to go without saying that she comes from the same supernatural source as the "promptitude merveilleuse" of the servants ("marvellous promptness", 322). Alvare's off-hand comment to Biondetto serves as the summons to the singer – an inference which readers can make if they adopt the "supernatural" super-prior to the situation. Then, the message has no "bizarrerie" for readers and the inference remains preconscious (or "unconscious" in Helmholtz' sense).

Fiorentina's "deshabillé" and her excuses about it, on the other hand, suggest something unexpected. If she is indeed summoned by the devil to appear before Alvare, why does she not know (or why did he not tell her) that Alvare has company? If we think further about this, however, we can explain this instance either through Alvare's delusion and vanity (because he thinks highly enough of himself to expect that a fêted opera singer will appear just for his own pleasure) or through the devil's well-judged play on Alvare's vanity (making it appear as if she just came for Alvare). The situation seems strange but it does not call for immediate disambiguation. Instead, it can serve readers as a reference point later on. For example, when Alvare describes his enchantment at Fiorentina's song, he says: "J'étais ému jusqu'au fond du cœur et j'oubliais presque que j'étais le créateur du charme qui me ravissait" (323).⁸ In the earlier instance, too, he might just as well have forgotten that his wish had been the creator of Fiorentina's appearance. As the narratorial discourse reminds readers of the supernatural powers of the character (ironically by stating that he had forgotten about them), they can revisit the earlier instance and inscribe it (more or less consciously) into the general predictive model which they are currently developing.

In *Le diable amoureux*, Cazotte sets in motion an intricate machinery that draws on "unconscious inference" (of the kind that pervades our everyday cognition), the conscious inferences of Alvare as narrator, and the (more or less conscious) revisiting of previous inferences on the part of the reader in light of new information. At times, the narrative contextualisation does not serve to create cognitive "penetrability" through which expectations shape percepts, but instead, creates the narrative, verbal equivalent of a duck-rabbit. In the dinner scene, both Alvare's act of vanity and the devil's manipulation of him remain a possibility when considered generally, but as soon as we start thinking through the implications of each of these options, we lose our grasp of the other option. Is Fiorentina in "deshabillé" due to the volition of Alvare (and his need to show off in front of his cabalist friends)? Is it due to the devil's psychological skill of granting Alvare his wish in such a way as pleases his vanity best and gives him the impression of being in control? These instances of ambiguity build up throughout the narrative. They constantly remind readers of the basic conflict between the realist hyper-prior (Alvare's delusion) and the supernatural hyper-prior (the devil's trickery), and they make it difficult to disambiguate between the competing hyper-priors, because the chains of inferences that these enable can be pursued to such a degree that we lose the other option from view.

* * *

What then, in such a wilfully ambiguous set-up, would be feasible disambiguation devices? Do they constitute something like a "found science" in verbal narrative? And what are their larger implications for the study of cognition?

⁸ In English: 'I was moved to the bottom of my heart and I almost forgot that I was the creator of the charms that delighted me.'

Cazotte's novella starts with a general statement from Alvare as the first-person narrator which orients readers in the time and space of the narrative, much in the tradition of the popular genre of the "mémoire": "J'étais à vingt-cinq ans capitaine aux gardes du roi de Naples" (315).⁹ The communicative situation that the narrator evokes takes readers back to an earlier stage of his life. It implies (1) that Alvare as the narrator can shape his narration of the events to whatever degree he likes, because he is our only source of its authenticity, and (2) that the narration takes place after the events, so that the narrating I of Alvare already knows how the story of the experiencing I of Alvare is going to end. Cazotte, however, decides to mix the features of the narrating I of Alvare with those of the experiencing I of Alvare. Throughout the novella, we get passages narrated in the present tense. For example, after Alvare has heard from the peasant Marcos that Biondetta has left, the narration turns to the present tense: "Marcos sort. Machinalement je me frotte les yeux" (372).¹⁰ He continues to wonder whether the seduction of Biondetta had actually taken place and is interrupted in these thoughts when his carriage is announced. "Je descends du lit; à peine puis-je me soutenir, mes jarrets plient sous moi" (372).¹¹ In these instances (as also in the description of Fiorentina's entrance quoted above), the narration becomes immediate, and it seems as if the narrating I of Alvare loses the distancing mode of his narration, just as the experiencing I of Alvare loses control over his limbs. Are these instances due to the devil's machinations that threaten to control the narrative (and can they hence be related to the supernatural hyper-prior)? Or does Alvare (temporarily and without acknowledging it explicitly) get shaken out of the flow of his delusions (and can this instance hence be related to the realist hyper-prior)?

The narrative of the novella itself ends with the words of Quebracuernos, a doctor of Salamanca. Considering, he says, that none of the strategies of the devil that we find in *Le diable amoureux* can be traced back to earlier demonological literature (such as Bodin's *Déonomanie* and Bekker's *Le Monde enchanté*), it seems likely that the devil has devised new strategies and is more dangerous than ever. The devil could come once more for Alvare, and the only thing that will keep him from falling into seduction would be a wife who has celestial qualities, such that "vous ne serez jamais tenté de la prendre pour le Diable" (376).¹² The doctor's own assessment of the situation keeps the ambiguity alive. Is it likely that Alvare will find a wife whom he would not mistake for the devil? This is presented, explicitly, as a strategy for warding off the devil's influence, but we wonder whether it is not Alvare's imagination (rather than the devil's machinations) that led him to take Biondetta for the devil. Indeed, the entire statement of Quebracuernos is reported by Alvare, so that readers cannot even be sure how reliable the words of the man of science are. In terms of a "found science" of the literary text, these instances of Quebracuernos' argument from authority and Alvare's potentially unreliable report of it keep the ambiguity between hyperpriors alive on the level of interpersonal cognition, where one might come to ask whether models that others teach or endorse are reliable or prone to prejudice (see Frith 167-183).

Even if the fictional text remains ambiguous to the very end, perhaps Cazotte himself can be prevailed upon to tell his readers how to read his novella? In the epilogue, he tells us of the different incarnations which his novella went through. In the version of 1772, Alvare sees through the tricks of Biondetta and escapes the snares of the devil. Cazotte then reports a second version (presented only to "personnes de sa connaissance"; "acquaintances" 377), in which Alvare falls prey to the devil and suffers the well-known consequences of eternal damnation. Finally, in the version of 1776, Cazotte seeks to combine both options, creating a narrative in which "Alvare y est dupe jusqu'à un certain point, mais sans être victime"

⁹ In English: At the age of twenty-five I was a captain in the guard of the king of Naples.

¹⁰ In English: Marcos leaves. Mechanically, I rub my eyes.

¹¹ In English: I get out of bed; I can barely support myself, my knees buckle.

¹² In English: you would never be tempted to take her for the Devil.

(377).¹³ Only this final version of the novella provides readers consistently with the kind of ambiguity that leads to the sustained hesitation of Todorov's fantastic. Strictly speaking, all three versions of Cazotte's novella would tend towards the marvellous, because the devil plays a role in the fictional world, but they do so to different degrees, because the powers of the devil change across versions. As Todorov acknowledges, "the fantastic in its pure state" is best "represented by a median line" between fantastic texts that tend toward the uncanny and those that tend toward the marvellous (44). Nevertheless, one can read the versions of Cazotte's novella as a process of experimentation, comparable to the artists' sketchbooks which Gombrich analyses in *Art and Illusion*. Smaller changes between the 1772 and the 1776 versions, and the long alternative ending of the 1776 version, presumably lead to vastly different effects on readers. Here, Cazotte's novella might offer a ready-made experimental design for the empirical study of reading and hence of predictive processing more generally.

A final duck-rabbit that enhances the ambiguity of *Le diable amoureux* is the very title of the novella. Who is the "devil in love"? At first glance, it seems most likely that this amorous demon is Biondetta. After the seemingly successful seduction, she reveals herself to be the devil (which – at this point – corresponds to readers' expectations), and yet at the same time, rather than triumph over the hapless soul she has snared, she confesses her love for Alvare once more ("ce cœur qui t'adore", 370; "this heart that adores you"). Even if characters make definite statements, confirming either the natural or the supernatural, Cazotte immediately supplies clues that make a conclusive inference problematic. Is the devil actually in love with Alvare? That assumption would work against any standard expectations as to the devil's character and his actions, and in turn might suggest that the entire narrative is created by Alvare's delusion. Franc Schuerewegen suggests an alternative interpretation of the title: none other than Alvare himself could be called the "devil in love". Schuerewegen traces throughout the narrative instances of Alvare copying Biondetta's behaviour, as well as making promises and declarations that serve to set her up (65): he begins to imitate the strategies of the devil (69) It seems to me an open question whether readers wonder about the title, which they encounter on every even page in the running head of the text (in a traditional edition), as they read the novella. Similarly, it seems to me an open question whether readers pick up on the similarities between Alvare's promises and ruses and those of Biondetta. What, in terms of "found science", is then the role of what we might call "super hyper-priors", such as titles or general tags of situation (like "doctor's visit" or "dissertation viva")? Surely, they provide predictive set-ups for our cognition in certain moments, but in how far are we (or do we need to be) aware of these tags for these predictions to take effect? Or in how far do we simply forget about them in order to facilitate cognition in the situation itself?

* * *

One of the earliest and most powerful claims of cognitive approaches to literature is that the human mind works through devices that are commonly considered "literary", such as metaphor (Lakoff and Johnson 1980; Turner 1996) and narrative (Bruner 1990; 1986). Fiction and the imagination, as scenarios we entertain and as a "default mode" of thinking, have more recently entered the fray (Richardson 2011). Arguably, the study of literary texts is as important as the cognitive sciences in the endeavour of working out the elements of the "literary mind", because it helps make these more or less automatic features of cognition noticeable and thus subject to analysis.

¹³ In English: "Alvare is the dupe to a certain point, but he does not become the victim." Cazotte goes on to add that the devil might have duped Alvare but, if he did, then Alvare would still have retained his virtue and hence triumphed over the devil (377).

In the predictive processing model we have virtual predictive models, or “fantasies”, which guide our perception and our cognition more generally. The literary genre of the fantastic highlights hesitation in the use of these “fantasies” in cognition, and hence it can serve as a repository of “found science”, of the ways in which predictive processing operates between different modes of cognitive “penetrability.”. Cognitive literary study, informed by Bayesian models of cognition and the literature of the fantastic, can contribute to studies of how we refine our predictive, probabilistic models and of whether awareness contributes to or detracts from these recalibrations. In the cognitive sciences, the literary strategies of the fantastic can then be employed for the design of experiments to study the cognitive penetrability entailed by more general, propositional knowledge of the situation, as well as the functions of awareness of such knowledge and its interpersonal reliability. In cognitive literary studies, we can pursue “found science” through the ways in which authors experiment with different effects in manuscript drafts or editions, how they make conscious the unconscious inferences that predictive processing depends on, and how literary history more generally provides us with a body of evidence for our “fantastic cognition”.

- Armstrong, Paul B. *How Literature Plays With the Brain: The Neuroscience of Reading and Art*. Baltimore: Johns Hopkins University Press, 2013.
- Baxandall, Michael. *Patterns of Intention: On the Historical Explanation of Pictures*. New Haven and London: Yale University Press, 1958.
- Bruner, Jerome. *Actual Minds, Possible Worlds*. Cambridge: Harvard University Press, 1986.
- Bruner, Jerome. *Acts of Meaning*. Cambridge: Harvard University Press, 1990.
- Cavanagh, Patrick. “The Artist as Neuroscientist.” *Nature* 434 (17 March 2005): 301-307.
- Cazotte, Jacques. “Le Diable Amoureux.” *Romanciers Du XVIIIe Siècle II*. Ed. Marguerite du Cheyron. Paris: Gallimard, 1965. 303-378.
- Clark, Andy. “Whatever Next? Predictive Brains, Situated Agents and the Future of Cognitive Science.” *Behavioral and Brain Sciences* 36.3 (2013): 181-204.
- Friston, Karl. “The Fantastic Organ. Book Review.” *Brain: A Journal of Neurology* 136 (2013): 1328-1332.
- Frith, Chris. *Making Up the Mind: How the Brain Creates Our Mental World*. Hoboken: Wiley Blackwell, 2009.
- Gadamer, Hans-Georg. *Wahrheit und Methode: Grundzüge einer Philosophischen Hermeneutik*. Tübingen: Mohr, 1960.
- Gombrich, E. H. *Art and Illusion: A Study in the Psychology of Pictorial Representation*. London: Phaidon, 2002 (1959).
- Hohwy, Jakob. *The Predictive Mind*. Oxford: Oxford University Press, 2013.
- Ingarden, Roman. *Vom Erkennen des Literarischen Kunstwerks*. Tübingen: Niemeyer, 1968.
- Iser, Wolfgang. *Der Akt des Lesens: Theorie Ästhetischer Wirkung*. München: Fink, 1976.
- Jensen, Melinda S. and Kyle E. Mathewson. “Simultaneous Perception of Both Interpretations of Ambiguous Figures” *Perception* 40 (2011): 1009-1011.
- Kandel, Eric R. *The Age Of Insight: The Quest to Understand the Unconscious in Art, Mind and Brain, From Vienna 1900 to The Present*. New York: Random House (2012).
- Kesner, Ladislav. “The Predictive Mind and the Experience of Visual Art Work.” *Frontiers in Psychology* 5: 1417 (2014): 1-12.
- Kukkonen, Karin. “Bayesian Narrative: Probability, Plot and the Shape of the Fictional World.” *Anglia* 132.4 (2014): 720-739.
- Lakoff, George and Mark Johnson. *Metaphors We Live By*. [Updated edition] Chicago: University of Chicago Press, 2003 (1980).

- O'Regan, J.K. and Alva Noë. "A Sensorimotor Account of Vision and Visual Consciousness." *Behavioral and Brain Sciences* 24.5 (2001): 939-973.
- Richardson, Alan. "Defaulting to Fiction: Neuroscience Rediscovered the Romantic Imagination." *Poetics Today* 32.4 (2011): 663-694.
- Ricœur, Paul. *Time and Narrative*. 3 vols. Transl. Kathleen McLoughlin and David Pellauer. Chicago: University of Chicago Press, 1984-1988.
- Schuerewegen, Franc. "Pragmatique et Fantastique Dans 'Le Diable Amoureux' de Cazotte." *Littérature* 60 (1975): 56-72.
- Starr, G. Gabrielle. *Feeling Beauty: The Neuroscience of Aesthetic Experience*. Cambridge, Mass: MIT Press, 2013.
- Todorov, Tzvetan. *The Fantastic: A Structural Approach To A Literary Genre*. Transl. Richard Howard. Ithaca: Cornell Paperbacks, 1975 (1970).
- Turner, Mark. *The Literary Mind*. Oxford: Oxford University Press, 1996.