



Social Understanding in Fictional Contexts and the Question of Error Detection

Joanna Klara Teske 
John Paul II Catholic University of Lublin
jteske@kul.pl

Arkadiusz Gut 
Nicolaus Copernicus University in Toruń
arkadiusz.gut@umk.pl

Received 29 December 2020; accepted 24 April 2021; published 16 September 2021

Abstract

The paper examines the possibility of using three theories of mindreading (social understanding) – Theory Theory, Simulation Theory and Interaction Theory – and, more precisely, the cognitive mechanisms each of them postulates to account for the reader’s understanding of fictional characters as well as their real authors. Interaction Theory, which explains social cognition in terms of direct, situated, both bodily and minded interactions of agents, might seem irrelevant in the situation from which one agent is, strictly speaking, absent and the other agent’s body does not seem involved. Yet the paper suggests that the interaction might take the form of the reader’s simulated, both minded and bodily, one-sided (pseudo)-interaction with characters, possibly as one of the characters, and artefact-mediated interaction with the real author. The paper suggests further that each theory can account for some error-detection mechanism which might help the reader to differentiate between correct and incorrect ascriptions of characters’ mental states and/or predictions of their next action and/or vicarious anticipations of the reader’s interactions with characters, thus improving the reader’s mindreading skills (social understanding). All these ideas are illustrated with analyses of the hypothetical reader’s mindreading response to passages taken from three works of fiction representing different conventions (mimetic and anti-mimetic, under- and over-representing mental experience, featuring many and few social interactions).

Keywords: social cognition; social understanding; narrative fiction; error detection; interaction theory

1. Introduction

Theory Theory (TT), Simulation Theory (ST) and Interaction Theory (IT) all strive to explain how people read each other's minds, how they understand each other. It has recently been argued that the theories might also be used to analyze the cognitive aspect of the reader's encounter with fictional beings (characters, narrators, implied authors) as well as (less directly) the real author who has called them into being. While taking into consideration all the three theories, the present paper aims to focus specifically on IT, whose relevance for literary contexts may appear doubtful. It further aims to inquire into the possibility of the reader's ability to detect error in her ascription of mental states to characters, predictions of their next actions and vicarious anticipations of interactions with characters. The paper thus extends our exploration of readerly mindreading from "The Reader's Mindreading of Realist, Modernist, and Postmodern Fiction: A Comparative Study", in which we focus on the effects of literary conventions, with their characteristic presentation of fictional mental experience, on the reader's ability to understand characters. While working on the same texts as before, we offer now a very close analysis of the reader's mindreading response so as to examine the error-detection possibility, which seems to be correlated, we argue, with the reader's gain in mindreading competence.

Social understanding with reference to fiction embraces the reader's relations with characters, the narrator, the narratee, the implied author,¹ and the real author.² Naturally, intratextual relations between two characters, the narrator and the narratee, or the narrator and a character can also be viewed in terms of social understanding. However, analyses of intratextual relations involving narrative agents may occasionally pose in this respect serious difficulties since the impersonal narrator, the impersonal narratee, the implied reader and the implied author, though they may be construed as fictional minds, lack one, two or even all of the following: (fictional) body, (fictional) personality, (fictional) ability to communicate verbally. Owing to space limitations, in the present text we investigate only two relations – the reader's relation with the character and the reader's relation with the real author.

The theoretical background of our project comprises various attempts undertaken recently by literary scholars and cognitivists to explore the mindreading activity of the reader. In particular, it has been argued that fiction provides the reader with a unique opportunity for mindreading practice. Informed about a character's behaviour, the reader infers the character's beliefs and desires using her personal "folk psychology" and predicts the character's next action in a way analogous to real-life situations. This argument, originally put forward by Zunshine, is based on TT, though Zunshine refers at times also to ST (2006, esp. pp. 3-44, 162-64). There are other theoretical possibilities. Palmer argues that the two theories – TT and ST – should be combined (2004, p. 144) and yet his work seems to be inspired first of all

¹ The narrative agent of topmost authority who in the text is the counterpart of the real – extratextual – author.

² The relation between the real reader and the implied reader – the reader modeled by the text – seems of little relevance in the context of mindreading as the real reader is more likely to identify with the implied reader than to interpret her or interact with her.

by IT. IT is also explicitly favoured by Herman (2011). In our previous publication we have argued that all three theories are complementary in their analysis of the reader's understanding of characters, even though IT, emphasizing as it does the moment of direct, embodied and situated interaction, may at first seem irrelevant to literary contexts.

The question of errors in the reader's mindreading and the possibility of correcting them seems vital and yet to the best of our knowledge the subject has not yet received an in-depth treatment. In her original publication Zunshine suggests that we (readers) compare our attributions with "what we assume could be the author's" attribution, which would mean that she agrees that some kind of confirmation of one's suppositions is desirable; but then Zunshine adds that the reader's mindreading satisfaction does not really depend on the correctness of his mindreading hypotheses: "from a cognitive perspective, a misinterpretation of a character's state of mind is still very much an interpretation, a fully realized and thus pleasurable engagement of our Theory of Mind" (2006, pp. 24-25). We would like to argue that for the experience of literary mindreading to be cognitively beneficial it is important that at least from time to time correct attributions be confirmed by the text and reinforced by the reader's sense of satisfaction and the incorrect attributions be disconfirmed.

In the paper we argue that (1) TT, ST and IT can help explain how readerly mindreading can help the reader improve her mindreading competence/ her social understanding skills because they show that the reader's mental ascriptions, predictions and enacted anticipations of interactions made with reference to characters can to some extent at least be followed, monitored and checked by error-detection control, and (2) IT can complement TT's and ST's description of the reader's cognitive response to the real author, though the interaction between the reader and the real author departs from the usual face-to-face forms of social understanding and no error-detection system can be found here unless the reader has access to relevant extra-textual information.

As in our previous project, we take into consideration three narrative styles – realism, modernism and postmodernism – and analyze heterodiegetic narratives: (narratives in which the narrator does not take part in the main narrative). As regards realism, we focus on the variety in which mental states are under-reported (in narratological terms this means the use of objective-realist narration, in which the narrator has no insight into characters' minds). In modernism, by contrast, mental states are typically over-reported (often by means of stream-of-consciousness techniques). While realism typically entails many passages featuring social interactions, modernism features extensive introspective passages showing the experience of isolated characters. Both realism and modernism are mimetic – they aim to represent either external or internal reality. The third style we discuss – postmodernism – abandons the principle of verisimilitude and exposes the text's artefactual character (by means of, among other things, diverse intertextual, self-contradictory or fantastic elements); it is, in other words, relatively anti-mimetic and metafictional. As regards over- and under-reporting of characters' mental states, as well as the social vs. individual character of human experience, postmodern fiction often strikes some kind of balance. We thus do not take into account the rich assortment of narrative techniques the three conventions have developed, for the sake of contrast focusing on their most characteristic features.

In Section 2 we briefly explain how the three theories describe a real-life social encounter, how they perceive sources of errors in mindreading and possibilities of their detection, and what research has been done on these phenomena. Next, in Section 3, we briefly explain how the theories might apply to the phenomenon of the reader's encounter with fictional characters and we cite some IT theorists on the possibility of mediated and one-way interactions to argue that the literary context is not the only irregular situation that IT can find challenging but not beyond its explanatory power. In Section 4 we analyze short passages exemplifying three narrative styles – realism, modernism and postmodernism – to show in the light of each of the three theories how the reader might respond to the fictional characters' and their authors' mental experience revealed or concealed in the text. We consider correct and incorrect mindreading responses to investigate whether the reader gets any feedback on the quality of her mindreading work. In Section 5 we try to formulate general hypotheses concerning (1) TT's, ST's and IT's treatment of error detection in fictional contexts with reference to characters, and (2) TT's, ST's and IT's accounts of the reader's interaction with the real author (the issue of the mindreading error detection included).

2. Introduction to TT, ST and IT: their interpretation of error and error detection in real-life mindreading

One of the most popular theories that seeks to explain the phenomenon of mindreading is called Theory Theory (TT). It defines mindreading as a complex ability to make sense of and reason about mental states, especially beliefs, intentions, and desires. TT examines how these states function in normal everyday explanations and predictions of other people's behavior. According to TT, the mind has the status of a theoretical construction, mental states of others are not directly observable; hence we attribute mental states to other people in the process of inference based on our observation of their behaviour and our (personalized) folk-psychology (FP) theory (attributions of mental states are taken as final hypotheses in an inferential process). Mindreading depends in large part on cognitive processes or mechanisms, the most important of which is, according to some authors, the Theory of Mind Mechanism (ToMM) – an innate automatic mechanism, dedicated to mindreading which spots signals crucial for attributions of mental states (Wellman, 2014). This approach, located in the tradition of cognitivism, is deeply rooted in the post-Cartesian tradition. It is a mentalistic approach (i.e. it uses mental terms in explanations), endorses internalism (sees mental states as internal), and views the mind as a symbol-manipulating inference “engine” which maps reality via representations. TT states further that on each level and at each developmental stage, mindreading entails the use of amodal concepts and quasi-theories. To recap, one may say that according to TT, mindreading other people is essentially a theory-driven activity.

Historically, the first alternative to TT was Simulation Theory (ST), which holds that mindreading is effected largely by the natural ability of humans to simulate (internally imitate) the experience of others. The simulation process consists of three stages: first “the attributor creates in herself pretend states intended to match those of the target”, next she processes them e.g. in “a decision-making or emotion-generating mechanism” and finally “assigns the output state to the target” (Goldman & Sripada, 2005, pp. 80-81). ST clearly differentiates

two levels of mindreading: high-level mindreading targets mental states of a relatively complex nature, such as propositional attitudes, and the process of simulation has some degree of accessibility to consciousness (Goldman, 2006, p. 147); on the low level mindreading involves the recognition of emotion types (cf. anger, fear, surprise) without identification of any propositional contents – it is “comparatively simple, primitive, automatic and [...] often based on the ‘mirror-neuron’ system” (Goldman, 2006, p. 133). Crucial for ST in the mindreading process is enactment imagination – imagination stimulated by emotional and bodily experience – and the mirror-neuron system mentioned above. Endorsing introspectionism, ST postulates that introspective mechanisms are used to read first-order (1st-person) mental states generated as a result of simulation, which are then projected onto the other person. To sum up, one may say that according to ST, mindreading is a process-driven activity, i.e. the simulating activity of the mindreader is driven by the mental process of the person whose mind is being read.

The interactive approaches going under the umbrella term IT, which constitute the theoretical alternative to TT and ST, include Gallagher’s interactionism (2001), Hutto and Myin’s radical enactivism (Hutto, 2008; Hutto & Myin, 2017), phenomenological approaches (Zahavi, 2001), the action-based perspective proposed by Bickhard (2009, 2013), autopoietic enactivism (De Jaegher & Di Paolo, 2007; Di Paolo, Cuffari, & De Jaegher, 2018), and embodied cognition (Thompson 2007). What all supporters of IT share is the idea that the mind emerges as a result of interactions involving the subject in the context of the social world. For IT, social understanding takes place in social interactions with a considerable use of our “capacity to perceive facial expressions, hand gestures, bodily postures, and goal-directed actions as meaningful embodiments of the other’s mind” (Froese & Gallagher 2012, pp. 446-447). Being engaged in social interaction is understood as occasion for generating affordances for actions. Actions and interactions, in turn, are seen, at least by some proponents of IT, as material for future reflection, which abstracts from them social schemes and action strategies and allows the emergence of a flexible social mind. (Those IT authors who recognize representations as a resource in the cognitive process see them as embodied, not as information encoded in the form of inner mappings of reality, as postulated by TT). Historically speaking, IT alludes to and is in part the heir of Vygotsky, Piaget, and late Wittgenstein (cf. his model of language) as well as of ecological psychology, dynamical systems theory, Merleau-Ponty’s phenomenology and American pragmatism.

Proponents of IT often believe that the basic forms of intersubjectivity (discussed above) need to be supplemented with narrative and communicative practices and competences or social normativity. Since they are relevant to the following discussion of literary mindreading we briefly summarize these ideas below.

According to Gallagher, narrative practices “allow us to implicitly frame the other person’s situated actions; that is, we tend to see others in terms of a story that makes sense out of their actions. [...] Secondly, we use narrative reflectively and explicitly to provide reasons for our own actions and for the actions of others” (2014, pp. 604-605). The use of narratives is thus dual – they provide a structure within which the other can be understood, and specific information that helps us understand the other’s intentions (2012, p. 371). Throughout our lives we are exposed to narratives, which in the long run gives us a rich narrative knowledge of

a particularistic kind: “That wealth of particularistic knowledge, which may include knowledge about the specific individual with whom I am interacting, provides an insight or a way into the other’s situation [...]” (2014, p. 606). To sum up, for Gallagher narrative competence is the ability to interpret interactive situations in terms of stories; it is particularistic (unlike FP in TT), used both with and without reflection, with reference to the other and oneself.

Ratcliffe, in turn, argues that social understanding usually takes place “within a shared context of established social practices, which incorporate social norms (what ‘one does’ in situation x) and standardised artefact functions (what ‘one does’ with artefact x)” (2006, p. 43). We are defined by “normativity” – “a world of standardised artefacts, normative institutions and ways of doing things” (2006, p. 44). Social normativity thus facilitates the process of social understanding by putting considerable constraints on what action on the part of the other one can anticipate.

Finally, Hutto, like Gallagher, believes that for an interactive event to be meaningful for the person engaged in it, it must be construed narratively (Hutto 2007a, pp. 3-4; Hutto 2007b). Narrative practice is an inherent component of direct social perception (Gallagher and Hutto 2008). One might also say that internalized narratives with which we have had contact create a system of expectations about how we or someone else will behave. Unlike Gallagher, Hutto explains that “*everyday* understanding of intentional action is itself an *essentially* narrative practice – i.e. [...] ‘folk psychological’ understanding *always* takes the form of constructing narratives” (emphasis added, 2007a, p. 3). Narrative practice, understood as story-telling, is taken as a means of gaining access to FP concepts. In other words, Hutto incorporates the concept of FP, arguing that FP is narrative and in social understanding priority should be given to story-telling, since social meaning is by and large narrative in character.

All the above supplements to IT might be perceived as a kind of bridge between IT and especially TT, while on the whole the three theories – TT, ST and IT – offer quite different explanations of social understanding. Clearly, it is IT – which claims that mental states are directly given, there is no need for any mindreading, and that cognition happens in an interactive situation (rather than in a cognitive subject) – that stands apart. However, some authors see them as complementary. There is, for example, a hypothesis that an interaction between people begins on a pre-conceptual level that consists in capturing the perceptually available social situation embodied both in people’s behaviour and the environment and later, on this basis, continues on a conceptual and/or narrative level. Thus ST and IT better explain the pre-conceptual stage of social understanding, while TT is more adequate as regards the conceptual stage (Metzinger, 2014; Quadt, 2015). A somewhat similar approach is proposed by Gallagher, who believes that a person who engages in a social interaction has first of all a sensory-motor, perceptual, nonconceptual understanding of the other’s embodied intentions; only then can she resort to Theory of Mind or simulation (Gallagher, 2001). Elsewhere Gallagher argues that only in some contexts does IT need to be complemented by TT and ST: “It may be the case that confronted with some strange or unaccountable behaviour I do try to understand the other person by running a simulation routine (or by appealing to theory). This is clearly the rare case, however” (2008, p. 176). Newen shares this approach

when arguing that “theory-based inferences are important when we are confronted with cases that we find strange or surprising, i.e., situations where we meet another person suffering from a mental disease which we know nothing about, or where the person belongs to a culture that is radically different from ours” (Newen 2015, p. 4, cf. Gopnik & Meltzoff 1996). In his Personal Model Theory, Newen also explains that if the two questions – (1) how people read other people’s mental states and (2) how this kind of knowledge is subsequently organized and managed by the cognitive system – are set apart, it turns out that ST together with IT’s theory of perception answer the former question, while TT and NPH answer the latter (Newen 2015), which is yet another attempt to reconcile the three theories.

We find the interpretations of mindreading which in one way or another combine TT, ST and IT highly significant in that they help understand deeper the complexity of the phenomenon. At the same time we recognize the argument of the pluralistic approaches that the three theories make incompatible assumptions as regards the nature of the human mind and cognition and therefore cannot be easily combined (cf. e.g., Fiebich, Gallagher, & Hutto, 2016). In our article we therefore do not claim that the theories can or should be integrated, though it seems to us that by naming specific mindreading mechanisms they help capture different aspects of literary mindreading and to that extent complement each other. In the following discussion of the reader’s engagement with fictional characters and real authors we discuss TT’s, ST’s and IT’s view on the matter separately, but thereby we intend, among other things, to show that to fully understand the reader’s mindreading response one may perhaps need to take advantage of the findings of all three theories (possibly disregarding their theoretical backgrounds).

When reading others’ minds, people can make errors. Some of them result from extraordinary situations for which the mindreader is not prepared, such as contact with people suffering from mental problems or belonging to different cultures. Some derive from the faulty functioning of mindreading in standard situations. Below we try to compare and contrast the ways in which the three theories define the basic sources of errors in such standard situations and envision the possibility of their detection.

According to TT, errors are connected either with the so-called conceptual deficit – a missing or flawed element in the mindreader’s FP (Gopnik & Wellman, 1995) – or a faulty application of the theory to another person’s mental state. (*Nota bene*, a theory can never match the specificity, originality and individuality of other people’s mental states: “Mistakes, errors, ignorance, and wrong ideas are the very stuff of our everyday psychology,” Wellman, 2014, p. 6). A mindreading error is signalled by an incompatibility between the prediction of the other’s behaviour and the other’s actual behaviour. The faulty prediction undermines the mental ascription, which itself cannot be tested and on which the prediction is based. A mindreading error may also be detected in communication with the other if the other faithfully presents his mental states (Wellman, 2014; Gopnik & Wellman, 2012).

According to ST, errors may consist in the faulty functioning of the capacity for “imaginative identification” (Stich & Nichols, 1992). The mindreader may wrongly “replicate, mimic, or impersonate the mental life of the agent” (Gallese & Goldman, 1998, p. 497). This may happen as a result of introspective errors (cf. the use of 1st-person experience in the simulative process). Alternatively, the mindreader may fail to exclude her own mental states and

project them on the other, thus making an egocentric error (Goldman, 2013, p. 35). For ST, as for TT, an error in mindreading is detected when there is an incompatibility between the prediction of the other's behaviour and the other's actual behaviour. An error may also be detected in communication with the other.

The account of error and error-detection proposed by IT is significantly different. Since, according to IT, mental content is inherent in the processes of interactions, errors do not appear on the level of isolated inference-based attributions or predictions. Given IT's assumptions, errors consist first of all in misperception of the other's mental states (i.e. they appear in the misidentification of behaviours in which these states are embodied) (Gallagher & Varga, 2014). Errors can consist further in misinterpretations of context, mistaken anticipation of behaviour (formed on the level of reflection), or erroneous decisions based on inappropriate resources (Gallagher 2008a). For IT, error is detected when the agent notes that interactions fail (i.e. when the interactive partners are unable to achieve their aims).

To sum up, the mindreading error, which for TT and ST consists first of all in mistaken attributions and predictions, and which is therefore epistemic in character and can be detected because predictions often are testable (incompatibility between the predicted and the actual behaviours signals an error), for IT consists in failed (ineffective) interactions and is detected in interactions; it is pragmatic in character. (The contrast is not as sharp as it might appear since in the final account within all approaches mindreading errors can only be detected by the agents in the process of their interaction with the world).

As regards empirical research projects on errors in mindreading, those which concern adults focus on errors which can appear in ToM resources, i.e. in the mindreading system itself or in its executive functions (the precise location of the error is in fact a matter of contention). In particular, Mitchell *et al.* (1996) focused on the realist bias, Birch and Bloom (2007) investigated the impact of one's own knowledge on one's belief attributions (cf. the curse of knowledge phenomenon), Linet *et al.* (2010) as well as Keysar *et al.* (2003) examined errors resulting from the need to suspend one's own perspective when reading another person's mind, while Apperly *et al.* (2010) inquired into errors caused by mindreading of complex mental states determined to a considerable extent by the circumstances. By and large these research projects are theoretically grounded in TT (rather than ST and IT). Most importantly, they do not pay much attention to error detection mechanisms as such.

3. IT in the context of works of fiction

As stated above, TT and ST have already been applied by scholars in the analysis of the reader's mindreading activity (cf. esp. the pioneering work of Zunshine). According to TT, the reader's mindreading competence is basically his ability to infer the characters' mental states (beliefs and desires), and to predict their next actions, on the basis of the textual presentation of the characters' behaviour, their current (fictional) social situation, their image presented so far and the reader's FP. ST sees the reader's mindreading competence as her ability to imagine/simulate the character's experience (emotions, sensations, intentions) on the basis of textual information, choose her (the reader's) next action and project this experience and decision onto the character. It is less obvious how IT, which sees social under-

standing in terms of interactions, can approach the reader's understanding of characters (pseudo-cognitive subjects) with whom the reader enters into a pseudo-interaction.

One way in which this pseudo-interaction could be construed is as a kind of simulated but bodily one-sided interaction. The reader might in a make-believe mode interact with fictional characters either as himself or impersonating another fictional character. This may sound strange but in the literature such ideas have already been voiced. Without developing this idea, Rembowska-Pluciennik actually suggests that this kind of interaction might be essential for narratives: "A narrative is not only a representation of events: it also offers the reader simulated interaction with speaking, perceiving and cooperating subjects" (2018, p. 147). Ratcliffe, in turn, discussing the experience of a person watching a climber (i.e. in a non-literary context), speaks of "one-sided bodily dialogue" and proposes that the observer is bodily active (one "makes slight movements in response, clenches one's fists, grimaces," 2006, p. 46). One might assume that a similar one-sided bodily dialogue might be part of the reader's experience of fiction. Ratcliffe further speculates that people often imagine their interaction with other people with whom they are not in face-to-face contact: "When one 'simulates' what another will do, one often plays the part of oneself and imagines one's interactions with them. One simulates one's own words and actions as much as theirs and is sometimes surprised by what one does. One constructs interpersonal narratives, which play a role in structuring one's interactions in actual cases. On other occasions, one may, in some sense and to some degree, adopt their perspective. However, in so doing, one still simulates interaction between them and other people" (2006, p. 46). Such simulated interaction "has a perceptual, bodily and affective quality" (2006, p. 47). Again, one might assume that the reader might enjoy similar simulated interactions with fictional characters.

But there is no consensus here. Referring to Dynamic Systems Modelling, Gallagher argues that "When the [computer-simulated] agent attempts to interact with a non-responsive 'partner' whose movements are merely played back from a recording of a previously highly successful encounter, the interaction fails to materialize. Individual actors do not achieve their action performance by utilizing internal computational mechanisms, such as 'social contingency detection modules'. Rather, their successful performance constitutively depends on dynamical properties involved in their mutual coupling with the other. The give and take, back and forth mutual process makes the action on either side what it is" (2012, p. 192). If this Dynamic Systems Modelling evidence is relevant, it does not seem possible to speak of interactions when only one agent is active. Notably, Gallagher's approach seems consistent with the definition of interaction given by De Jaegher *et al.*: "Interaction: a mutually engaged co-regulated coupling between at least two autonomous agents where the co-regulation and the coupling mutually affect each other, constituting a self-sustaining organization in the domain of relational dynamics" (De Jaegher *et al.* 2010, p. 442)³.

³ The definition has recently been criticized. But even in the minimalist version suggested by the critics – "co-regulated coupling between conscious human beings," where the co-regulated coupling entails that "the actors knowingly affect each other's actions" – the basic idea that both sides of the interaction must be responsive stays (Schönherr and Westra 2019, p. 34).

In this situation we very tentatively propose that a kind of make-believe, one-sided interaction could take place between the reader and fictional characters, that might best be accounted for by a combination of IT and ST (or mechanisms they postulate). The reader's understanding of characters' intentions and actions can be further explained by IT with reference to narrative and communicative competencies, narrative practice or social norms.

Before we move on to the case studies, let us briefly explain how we construe the reader's mindreading activity in the process of her temporal reception of the text. Given that the reader can by and large control the pace of reading, choosing when to turn the page, we assume that the reader has time both to form mental ascriptions, action predictions or anticipated interactions before the further text reveals the authorial version of the current mental experience as well as subsequent behaviour of the character.

Finally, a terminological note is due. In literary contexts we speak of falsification/verification as well as confirmation/disconfirmation: the former are conclusive, the latter are not. Though falsification and verification in most contexts of human experience are out of human reach, in the context of fiction, if the authorial narrator describes a certain course of events and this presentation does not involve character focalization, such a description in most conventions is not subject to debate and hence may provide a falsifier or verifier, unavailable under normal circumstances.

4. The reader's socio-cognitive experience evoked by passages taken from Graham Greene's "The Destroyers", Virginia Woolf's "Mrs Dalloway in Bond Street" and Salman Rushdie's "Yorick", and its TT, ST, and IT interpretations: case studies

The following three case studies are meant to show in detail how the three theories presented above might describe the reader's activity as regards her understanding of fictional characters and their real author, when exposed to various kinds of narrative texts. The first passage comes from Graham Greene's "The Destroyers". Written in the realist convention, the story tells of a gang of boys in post-war Britain who demolish the house of Mr Thomas while the owner is away. At one point T., the boy who proposes this idea and who, once it is accepted, becomes the gang's leader, tells Blackie, the previous leader, that he (T.) has found some money.

"Did you find anything special?" Blackie asked.

T. nodded. "Come over here," he said, "and look." Out of both pockets he drew bundles of pound notes. "Old Misery's savings," he said. "Mike ripped out the mattress, but he missed them."

"What are you going to do? Share them?"

"We aren't thieves," T. said. "Nobody's going to steal anything from this house. I kept these for you and me – a celebration." He knelt down on the floor and counted them out – there were seventy in all. "We'll burn them," he said, "one by one," and taking it in turns they held a note upward and lit the top corner, so that the flame burnt slowly toward their fingers. The gray ash floated above them and fell on their heads like age. "I'd like to see Old Misery's face when we are through," T. said.

“You hate him a lot?” Blackie asked.

“Of course I don’t hate him,” T. said. “There’d be no fun if I hated him.”

Below we sketch the ways in which the three theories might handle the reader’s understanding of Blackie in this episode and the reader’s understanding of the real author in the whole short story. Please note that when discussing the reader’s hypothetical reaction, we do not argue that these are the reader’s conscious experiences.

Presumably the reader of the passage in her mindreading will focus on Blackie: Blackie’s mind is in the text disclosed most extensively; he is one of the most important individual characters (next to T.); and in some passages he acts as the focalizer.⁴ However, within TT, it makes sense to suppose that the reader can simultaneously mindread many characters – not only Blackie but also, for example, T. (cf. Oatley’s argument that in contact with fiction the reader practices reading many minds interacting with each other, 2016, p. 626). This possibility will be less obvious within ST and IT, which speak of simulation or simulated interaction, as both of these seem to demand greater concentration on one character.

Since the story does not contradict the reader’s knowledge of the world at any point (the principle of verisimilitude is respected in the passage), the reader will find it easy to use her FP beliefs; the reader’s mindreading competence will be activated in a fairly standard (typical of real life) way. Blackie’s mental states are not presented explicitly; there is no description of Blackie’s body to help the reader ascribe to him mental states, either. The context tells the reader that Blackie and T. are rivals – they are the former and the current leaders of the gang; T. is in charge of the destruction of Mr Thomas’s house – Blackie and the other members of the gang carry out his commands. In the episode in question T. singles Blackie out for the “celebration” – the act of burning the money. Supposedly the reader’s FP entails the following (or similar) beliefs: a displaced leader probably feels humiliated and might harbour a grudge; a member of a community who is shown favour by its leader feels privileged and proud; powerful people whose intentions are enigmatic should be viewed as potentially dangerous; people who feel confused often choose to be passive. On the basis of the contextual information and her FP beliefs, the reader may infer that Blackie feels confused: both distinguished and humiliated by his leader – this is the moment of mental ascription. She might next speculate that Blackie will not take any initiative – this is the moment of action prediction.

As regards the possibility of error detection (feedback on the reader’s mental attributions and predictions), the prediction that Blackie will be passive finds verification in the text: throughout the scene Blackie only asks three questions and does not undertake any actions on his own initiative. This verification of the prediction indirectly confirms the earlier ascription of mental state: confusion. If, on the other hand, the reader ascribed to Blackie the men-

⁴ This last function is possibly the most important in the context of mindreading, as it is the focalizer who elicits the reader’s empathy: “A character who is also a focalizer has a special claim not only on readers’ attention but also on their sympathy, for, as Bal observes, in such a case ‘the reader watches with the character’s eyes and will, in principle, be inclined to accept the vision presented by that character’ (Bal, 1985, p. 104).” (O’Neill, 1994, p. 96).

tal state of outrage and disbelief at the idea of the destruction of the money and predicted that the boy would try to hide at least some banknotes, the story's further course of events would falsify these hypotheses – the direct falsification of the prediction would indirectly entail disconfirmation of the earlier ascription.

A simulation theorist will approach the situation differently. There is little information about Blackie's looks and activity (he asks three questions and burns the banknotes), hence little input for mirror neurons to do their work and simulate on the low-level Blackie's mental state for the reader. But the high-level simulation performed by enactment imagination can take place. On the basis of contextual information the reader may put himself in Blackie's position, and then vicariously bear a grudge, feel pride and uncertainty. He can next process this insight and conclude that under the circumstances it is probably wisest to let things happen. The reader can then project these experiences and the resolution based on them onto Blackie. The reader might also make use of the poetic metaphor supplied by the narrator about the ash falling like age on the boys' heads – conjuring up a (possibly misplaced) feeling of untimely maturity, though it probably will not make much difference as regards action prediction at this point.⁵

As regards error detection, the story will verify the prediction of Blackie's passivity, and thus indirectly confirm the ascription on which the prediction is based. However, a thrifty reader, when simulating Blackie's experience, might commit the egocentric error and choose to hide some banknotes or challenge T. This decision, projected on Blackie as action prediction, will be falsified by the text. The reader's mindreading reaction either complies or clashes with the subsequent text, either way providing the reader with the relevant feedback.

The following presentation of IT's treatment of the reader's interactive response is highly tentative. In the analysis we do not take into account implicit presuppositions which, according to IT, are very much involved in the process of interactions. Acquired in the process of life, including enculturation, they constitute in a sense one's knowledge of the world (they define the world) and allow formulating anticipations. However, since presuppositions are unconscious (their content is not explicit), it would be difficult to reconstruct them. Presumably, the reader will find in the scene an affordance to pretend interact with T. as Blackie, simulating first Blackie's experience.⁶ The text offers the boys' verbal exchange and cooperative act of burning the money. In the verbal exchange Blackie merely asks three questions; two of his suppositions T. rejects (they will not steal the money, T. does not hate Mr Thomas). As Blackie, the reader may experience imaginatively what it is like to be out of one's depth in a tricky social situation. T. tells Blackie to come over, which Blackie and imaginatively the reader presumably do, then he kneels down, counts the money and lights the first banknote. The reader qua Blackie may at this point anticipate T.'s next action. He can anticipate that T. will indeed burn all the banknotes. Subsequently, the boys take turns

⁵ The use of metaphor, which goes beyond real life mindreading, can however stimulate the reader's imagination and enrich the simulative experience.

⁶ Within IT, it also seems reasonable to assume that the reader can imaginatively interact with Blackie and T. on his own behalf, without identifying with any of the characters.

and each burns thirty-five pounds. Presumably Blackie's need to be T.'s close companion is both granted (they burn the money together) and frustrated (T.'s answers probably estrange Blackie). The reader's reaction may well be bodily – he may for example strain his muscles as if adopting the kneeling position or become flustered by the criminal act. The understanding of T.'s action for the reader who experiences it as Blackie will develop in the process of the make-believe minded and bodily interaction. Additionally, IT explaining the reader's reaction might refer to narrative competency and the like. The reader might identify the meaning of the scene by placing the characters' actions in narrative contexts with which he is familiar – the context of boys showing off to impress each other when competing for leadership. Such placement will not involve mindreading or deduction, but will give the reader direct insight into the meaning of the scene. However, in the story under discussion, where we have two boys destroying the house of a man who has done them no harm, the situation is far from normative and this strategy may not be effective.

Error correction for IT consists in the make-believe interaction matching or failing to match the development of the interaction scripted by the story. The reader's enacted anticipation that T. will burn all the banknotes is verified by the text. If, however, the reader (qua Blackie), misreading the context as robbery rather than an act of pure destruction, made another anticipation, namely that T. would never burn any notes, his anticipation would be falsified by the story. So too would be the anticipation of the reader's own (qua Blackie) action, namely that he would try to persuade T. that they should give the money back to Mr Thomas.⁷ In real life for IT the measure of success is pragmatic and consists in the interactive engagement running smoothly. Successful interaction seems also to be the criterion of good social understanding for IT in literary contexts, when the interaction is a make-believe one. If the reader's imaginatively enacted anticipations comply with the text, the make-believe interaction runs smoothly and the reader's interactive skills are reinforced. Analogically, the reader's errors will be revealed by conflicts between the reader's and the text's perception of affordances for action, providing the reader with a chance to improve his social skills.

As regards the reader's attempt to understand Graham Greene's mind, TT and ST will not have much to offer. In particular, TT might argue that the reader's FP includes the belief that some people communicate by means of literary fiction so it is reasonable to search for a message in the text. Presumably TT will focus on the reader's attempt to identify the real author's beliefs behind the project (e.g. human evil is a mystery) and intentions (e.g. to share one's anxiety with others) relying on his FP. ST will argue that the reader can imaginatively put himself in the position of the real author – vicariously experience the need to write about juvenile delinquents. ST might thus attach much importance to the author's experiential situation reflected in the tale. Neither TT nor ST will provide much guidance as regards identifying the precise meaning of the communicative act. Since the situation does not involve face-to-face contact (the author of the story is long dead), IT will not be able to view

⁷ We realize that these anticipations are coarse-grained for a typical IT interaction (and thus misleadingly resemble TT's and ST's predictions), but Greene's text does not offer sufficient details for more fine-grained anticipations.

the situation in terms of basic intersubjectivity (direct reading of the other's intentions in the other's body language or intentional actions). But IT might make use of the narrative competency and argue that the reader might realize how there is a socially established narrative practice of writing fiction in which people engage if they want to communicate their ideas to a large number of people whom they do not know in person. This kind of understanding, though it can make Greene's act of writing the story comprehensible, does not help in the reading of the specific intentions or ideas he had in mind when writing the text either. To do this, one needs to resort to methods of literary interpretation and relevant cultural background knowledge. Once the reader has grasped the text's meaning, he can form some hypotheses about its author's mind. Significantly, unless the reader consults some extra-textual materials – critical reviews, interviews given by the author, biographical materials – he will have no chance of detecting an error in his interpretation. And the error, if it is there, will originate in the reader's interpretive, not mindreading, competence.⁸

The second passage is taken from a modernist short story by Virginia Woolf titled "Mrs Dalloway in Bond Street" and features the experience of Mrs Clarissa Dalloway, whose mind, while she is buying gloves, is preoccupied with all kinds of ideas concerning the shop-assistant's holidays, Mrs Dalloway's husband's (Dick's) opinions on charity, poetry presenting death as a well-deserved respite, and atheist belief:

"Madame's hands are so slender," said the shop-girl, drawing the glove firmly, smoothly, down over her rings. And Clarissa looked at her arm in the looking-glass. The glove hardly came to the elbow. Were there others half an inch longer? Still it seemed tiresome to bother her perhaps the one day in the month, thought Clarissa, when it's an agony to stand. "Oh, don't bother," she said. But the gloves were brought.

"Don't you get fearfully tired," she said in her charming voice, "standing? When d'you get your holiday?"

"In September, Madame, when we're not so busy."

When we're in the country thought Clarissa. Or shooting. She has a fortnight at Brighton. In some stuffy lodging. The landlady takes the sugar. Nothing would be easier than to send her to Mrs Lumley's right in the country (and it was on the tip of her tongue). But then she remembered how on their honeymoon Dick had shown her the folly of giving impulsively. It was much more important, he said, to get trade with China. Of course he was right. And she could feel the girl wouldn't like to be given things. There she was in her place. So was Dick. Selling gloves was her job. She had her own sorrows quite separate, "and now can never mourn, can never mourn," the words ran in her head. "From the contagion of the world's slow stain," thought Clarissa holding her arm stiff, for there are

⁸ Dutton in his *Art Instinct* argues that the current resignation from reading the author's intentions in literary studies is a mistake in the light of naturalist theory of art's biological function, which says that artefacts function as fitness-indicators (2009, pp. 167-177). The current discussion shows that a similar argument can be constructed with reference to cognitivist exploration of literary mindreading. To focus in scholarly interpretations of fiction entirely on the reader's needs, cultural contexts or the work itself seems a mindreading error on a macro-scale, a systemic failure to mindread the real author.

moments when it seems utterly futile (the glove was drawn off leaving her arm flecked with powder) – simply one doesn't believe, thought Clarissa, any more in God.

TT will not have much to say about the reader's standard mindreading activity (though such activity is encouraged by the text's mimeticism) as the reader has limited opportunity to read any mental states using his FP: that Clarissa is upset about the gloves which are too short, empathetic towards the shop-girl who might be having her period and doubtful about the existence of God is by and large explicitly shown. The reader may, however, speculate about the reasons why Clarissa believes that the shop-girl is tired, and deduce, using his FP (in particular the belief that people often project on others the sense of fatigue that they feel themselves) that Clarissa herself is tired. He may also wonder why Clarissa finds it so easy to justify her own passivity towards another woman whom she presumes to be miserable, and explain this behaviour referring to his FP belief that some women depend on their husbands for their social activity or his FP belief that some people indulge in fantasizing about their acts of charity. The reader might also try to predict Clarissa's next action – how she will try to be nice to the shop-girl but, being herself tired and worried, will eventually lose patience and leave the shop without saying goodbye. This hypothesis will be neither confirmed nor disconfirmed by the text, which never shows Clarissa leaving the shop. Thus, in theory, ascription of some deeper mental states (underlying those that are explicitly shown) as well as some predictions along the TT mindreading model might be feasible. Some predictions might be testable. The problem is that nothing much happens in the story, Mrs Dalloway (as is typical of modernist fiction) is lost in her own mind. The reader might perhaps predict the further development of Mrs Dalloway's thoughts and some such predictions of mental states might be verified or falsified by the text (e.g. the reader might predict that as time goes by, the protagonist will become more and more desperate only to realize that in Woolf's story she becomes more and more resolved to keep upright). Alternatively, the reader might mindread the shop-assistant though the text does not encourage this choice, the shop-assistant being a flat and inconspicuous character, who does not act as a focalizer at any point. Whatever desires and beliefs the reader chooses to ascribe to her will be neither verified nor falsified as the text never reveals the girl's mental states (though if the reader ascribes to the girl a state of fury, then her quiet demeanor till the end of the scene will indirectly question this ascription). The error-detection possibility is thus analogous to the case of Greene's text, though Woolf's reader has much less opportunity to try and read the characters' minds (as the mind of the protagonist is transparent) or check her action predictions (as the protagonist practically does not act). However, the reader can infer the deep mental experience hidden behind the more superficial one or the future mental experience from the present one (and some such inferences might be testable).

ST will explain the reader's mindreading of Clarissa in terms of simulation. The simulation will involve the reader's enactment imagination to a greater extent than mirror neurons, given the shortage of relevant descriptions of Clarissa's body language. The reader can simulate mental states which are presented by the text and those which can be reconstructed on the basis of their bodily traces and contextual information. Thus the reader can use her mind as a model for simulating Clarissa's belief that the gloves she is trying on are too short or her supposition that the shop-assistant might be having her period, which are clearly stated. On the other hand, Clarissa's being weary with life is merely hinted at. If the reader tries to

experience the snatches of poetry that recur in Clarissa's inner speech, he may feel troubled by the thought of death in a godless universe, or the thought that life means constant challenges from which only death brings relief. The reader might simulate Clarissa's fatigue and strain, prompted by the words "bother", "tiresome", "agony", "tired" recurring in her inner speech, and predict that Clarissa might become irritable if her desire for longer gloves is further frustrated. He might also listen to the melancholic rhythm of the final sentence and vicariously experience a sense of emptiness. It is not obvious how this experience of being troubled, tired and frustrated can be "translated" into action prediction of the character who is resolved to "keep upright" (as Mrs Dalloway is throughout the text; *nota bene* this resolution is missing from the passage cited above). Presumably the reader will project onto Clarissa the experience of hiding the turmoil of one's soul from other people. There is no significant action in the story anyway – i.e. things happen in Clarissa's mind (the standard definition of action in narrative theory does not treat mental events as events proper). Under the circumstances, the reader's simulation could entail simulating present deeper and predicting future mental states. As regards error detection possibilities, these are analogical here to the TT's account, i.e. considerably reduced in comparison with those accessible to the reader of Greene's text, available mainly in the realm of predictions of the character's future mental states.

Finally, IT will consider the reader's interaction with the text, taking the form of imaginative pseudo-interactions as a character (or oneself) with other characters. The text features an interaction between Mrs Dalloway and the shop-girl. The reader identifying with Mrs Dalloway (the focalizer and the protagonist) may, in a sense, feel the glove being firmly pulled over his outstretched hand. He may also feel pleased with having slender hands and embarrassed with having to make yet another request. On closer inspection, however it turns out that the interaction in question only marginally involves Clarissa's bodily and social experience (in the passage, information about the real world in contrast with Clarissa's inner experience is twice placed in brackets). By and large Clarissa is lost in her thoughts: she imagines the shop assistant's period, her holiday and her being reconciled with her social position (the text does not supply any "external" evidence supporting her fantasies); she remembers fragments of mournful poetry and ponders on God's absence. Because her interaction with the shop-assistant is highly conventional (a client and a shop-assistant performing their roles) and at the same time Clarissa has very poor contact with reality (mostly she is daydreaming), the reader's experience of Clarissa's face-to-face interactions will be minimal. The reader may vicariously anticipate Clarissa's actions, e.g. that Clarissa will realize how her husband's beliefs need not stop her from being generous – after all, death ends everything for good in a godless universe so little harm will be done if the shop-assistant gets a proper holiday. The anticipation will conflict with the text, in which Clarissa never comes to this realization. The discrepancy between the reader's imagination and the fictional text might work as a kind of error indicator. Alternatively IT might argue that the reader, making use of his communicative competency and assumptions of naive normativity, will understand the interaction between Clarissa and the shop assistant as being regulated by the common script of the customer served in a shop. However, this strategy once again will not be effective since Clarissa behaves in an absent-minded way, without really engaging in her interaction with the shop-assistant.

As regards the reader's attempt to understand the real author's mind interpreted by the three theories – TT, ST and IT – one can first of all repeat that, as in the case of Greene's story, acquaintance with literature and its artistic means of expression (not covered by the standard mindreading competence as interpreted by TT and ST or IT) will be crucial; and the kind of artefact-mediated interaction will not comply with the typical (direct) interactions (assumed by IT). But taking advantage of TT, one might say that the reader, using her FP (in particular the belief that verbal statements serve communicative purposes), can infer that the short story reflects Woolf's desire to share her ideas with the reader. Delving deeper into the meaning of the narrative and using more specific FP resources (in particular the belief that people usually communicate ideas which appear vital to them), the reader might infer that Woolf is interested in the experience of women or the complexity of the human mind. ST might explain these insights into Woolf's mind by suggesting that in his imagination the reader sees himself as the author of the short story and experiences a communicative need as well as, more specifically, the author's beliefs about the mind's complexity (also manifest in the reader's difficulty grasping what is going on in the protagonist's mind) or feelings of apprehension related to the theme of death. IT in turn might perceive the reader's attempt to enter into an artefact-mediated interaction with the real author in terms of her communicative and narrative competency. Such interaction would involve inquiring first of all into the author's intentions in the process of the reader's embodied experience of the text. This however is problematic. The problem is not only that the interaction is mediated but also that the author does not address any particular reader and that the reader cannot communicate with the author, Woolf being dead for many years. Any interaction the reader and the author might under the circumstances have will be fairly reduced. As regards the possibility of the reader monitoring her understanding of Woolf's mind (based on her understanding of the text), the situation is parallel to that with Greene's mind. If the reader's interpretation of Woolf's emotions and ideas is faulty, the fault is connected mainly with the reader's interpretive, not mindreading, skills, and can apparently be identified only by reference to extra-textual materials.

The last passage exemplifies postmodern fiction and comes from "Yorick" by Salman Rushdie. The text retells the story of Hamlet, making Yorick the protagonist. The passage comes after the incident in which Hamlet, misinterpreting his father's (Horwendillus's) amorous advances towards his mother, Queen Gertrude, as murderous intentions, attacks him and is subsequently given a severe beating. The narrator asks the "reader" to speculate about Hamlet's mental state, then describes this state at great length, in a hypothetical mode, and playfully ascribes these hypotheses to the "reader":

Hamlet Alone: (But I'll leave soliloquies to better pens; my vellum's silent on what Hamlet felt while locked and wealy in his room – you must infer his thoughts from what he did. – It's possible you may see him haunted; a Horwendillian phantom shimmers before his eyes, and seems to miss no chance of squeezing the life-breath from the Queen. – Amlethus' eyes, made visionary by fear, observe the frenzied spectre as it assassinates Queen Gertrude, oh, a hundred, no, more, a thousand times, [...]. until, in short, he (Hamlet) sees his recent rescue's no end, but only a beginning, to his loving anguish; and appreciates the necessity of finding a permanent solution to this matter. – So a plot is born, conceived by nightmare urgency out of hate, its generative organ that royal whip that stung his royal

buttocks, delivering upon those nether cheeks just such a yoricking as he'd often given the fool. [...] – Now, reader, my congratulations; your fancy, from which all these dark suppositions have issued (for I began by swearing myself to silence), is proved by them more fertile and convincing than my own.)

With reference to Hamlet and the reader's attempt to understand him, TT might make use of the reader's FP belief that a person who is punished and rejected may feel angry and rebellious, which is confirmed first in Hamlet's confused thoughts and then his decision to take revenge. If the reader's FP includes Freud's theory of the Oedipus complex, the reader might support her mindreading with this theory (Hamlet's resolution to kill his father will then be motivated originally by his own desire for his mother). Similarly within an ST account, the reader might imagine how Hamlet feels struck by the thought that his father intends to murder his mother, with his body still painful after the beating, and predict that now he in turn might resort to violence. IT might put in the centre the reader's simulation of Hamlet's interactions with his fantasies about his father's murder of his mother; or speak of the reader's use of his narrative competency to make sense of the story of the son's rivalry with his father, on the condition that a script of the kind is for the reader normative.

But the reader may just as well be discouraged from engaging in such mindreading activity by the clear signal that instead of Hamlet's experience she is offered allegedly her own speculations concerning this experience – speculations which are ridiculed as naive parodies of Freud's interpretation of the Oedipus complex. A discouraged reader may adopt a passive attitude towards the text or engage in speculations about its author's reasons for this provocation. If the reader's mindreading activity with reference to the characters is indeed thwarted, there is little space for mindreading, errors and error-detection mechanisms.

As far as the reader's attempt to understand Rushdie is concerned, the situation is analogous to those involving Greene and Woolf though (1) the reader, confronted by a highly unusual text, will this time naturally direct her attention towards the author, not the characters; and (2) Rushdie is still alive. Since the reader is deprived of a chance to observe the author or interact with him in a normal real-life face-to-face situation, TT, ST and IT all seem to be at a disadvantage. Perhaps ST, referring to the reader's enactment imagination, might once again best explain the reader's guess that behind the playful, ironic and ambiguous text, there hides a troubled and defiant mind, but reading another mind on the basis of a verbal artefact created by that mind is a highly unusual kind of mindreading. Any attempt to read specific beliefs and intentions held by Rushdie when working on "Yorick" will require substantial knowledge of literature and culture and experience in literary interpretation. No error-detection mechanism seems available within any of the three theories, TT, ST and IT, unless some extratextual evidence is brought into consideration, including a real-life conversation the reader of Rushdie's text might have with its real author.

5. Theories of social understanding and fiction (fictional characters and their real authors) – discussion

In this section of our article we discuss the case studies presented above. The discussion is divided into two sections: one concerning the reader's understanding of fictional characters and the other concerning the reader's understanding of the real author. In the former we focus on the possibility of error-detection. In the latter we also sum up the findings of our modeling of the mindreader's work (this section is missing in the case of the reader's engagement with characters for we have already discussed this issue at length in our previous publication).

We argue that hypothetically speaking each theory – TT, ST and IT – can make sense (each in its own way) of literary mindreading with reference to fictional characters of narratives of various kinds, and that the theories, or the basic mindreading mechanisms they postulate, should best be taken together as each seems to capture some aspects of the phenomenon that the other two miss. Further, we claim that the impossibility of checking the reader's ascriptions of current (or deeper) mental states, predictions of characters' future actions and make-believe interactions with characters presumably limits the benefits of literary mindreading practice, but that, compared with real life, the possibilities of testing one's mindreading work in contact with literature are considerable. The reinforcement of one's mindreading work after a correct attribution/prediction or a smooth interaction presumably helps improve one's competence. As regards the reader's mindreading of the real author, there is little gain here as the reading of the text, regardless of the type of the narrative, is based by and large on the reader's literary competence and it is this competence that is mostly exercised. What is more, there is hardly any chance for error detection in this mindreading experience.

5.1. The reader and fictional characters – the possibility of error detection in the reader's mindreading work

The correctness of the reader's ascriptions of mental states to a character and predictions of her next action, for TT and ST, or the reader's anticipations of the character's next act inscribed in make-believe interactions, for IT, may be checked if the text, having presented the character's behaviour, reveals the character's actual mental state and next action. In realist fiction the former may but need not happen and the latter usually does happen. *Nota bene*, for TT and ST a verification of a prediction can be taken as an indirect confirmation of the ascription on which a given prediction is based if the mental state is never directly presented (if the state is presented, this presentation serves as verification or falsification of the ascription). In modernist narratives, if the character's next actions are revealed (often they are not, or if they are, they have little significance), the reader's predictions (or vicariously enacted anticipations) formulated on the basis of the text's presentation of mental states can be verified or falsified but ascriptions of mental states and correction of errors committed in the process will be rarer since the text usually from the start explicitly presents mental states of the protagonist limiting such mindreading activity. In postmodern narratives, falsification or verification of mental-state ascriptions and next-action predictions/anticipations will be available along the same scheme. Still, the postmodern reader will often find that her guess-

work conflicts with the text (or its author). FP-based ascriptions, simulations based on the reader's own experience and simulated interactions based on prior real-life interactions or narrative practice may more often fail as the principle of verisimilitude does not obtain in this kind of fiction. To sum up, as regards error detection, realism works best, while in modernism there are fewer opportunities for mindreading and in postmodernism probably not enough life-likeness. (For a comparison and contrast between error detection in literary and real-life mindreading, see the conclusions).

5.2. The reader and the real author – the possibility of error detection in the reader's mindreading work

The mind of the real author within the cognitivist framework seems of crucial importance: ultimately it is the real author whose mind the reader indirectly encounters when reading fiction, though the reader's attention may often be drawn in the first place to the minds of characters or narrators. However, TT, ST and IT seem to be able to offer only a limited account of the reader's attempt to understand the author's mind. To understand the text (and the real author's mental states expressed therein) the reader's literary competence (including knowledge of literary means of expression, of relevant literary traditions and cultural contexts etc.) seems of utmost importance. The differences between the three theories will concern mainly their justification of the reader's belief that a work of literature is a vehicle for communication or an element of mediated interaction.

TT can try to explain the reader's belief that the real author who produces a work of literature is engaged in communicative activity with reference to FP. FP will also be seen by TT as the reader's main resource when interpreting the text and ascribing specific beliefs and desires to the real author. In the light of TT, the reader who tries to read the real author's mind is deprived of the standard input for mindreading inferences (the other person's body language, specific social context); the narrative text taken as the author's verbal behaviour is highly untypical of real-life situations.

Within ST the situation is analogical. There is little possibility for the usual simulative experience based on observing the other, which normally triggers the mindreading process. But the reader can first of all simulate in her mind the mental state of someone engaged in the act of writing a piece of literature and next, reading the text, may try to simulate the specific experience of the real author communicating the specific thematic content, and project it onto the author.

Neither TT nor ST will have much to say about the reader's prediction of the real author's next action, even if the reader can arrive at a mental ascription based on the reader's literary competence. The lack of action prediction is related to the lack of situational context (the reader has usually no idea about the real author's life situation when writing the text), the author's relative anonymity (the reader knows the author through the prism of the work but this is a highly fragmentary knowledge), and the work's equivocacy (many works convey multiple and ambiguous meanings and thus conjectures about the author's beliefs or sensitivity based on the text's interpretation remain uncertain).

Defining the reader's cognitive relationship to the real author, IT can speak of mediated social understanding in which the moment of interaction is greatly reduced: in principle the real author's creation of the work marks the beginning of the interaction and the reader's act of its reception marks its completion (unless the reader gets in touch with the author). The form of this interaction and its course differ very much from everyday face-to-face interactions, although in the age of advanced technology and the epidemic a huge number of human interactions are mediated in one way or another. *Nota bene*, in the case of deceased authors (like Greene or Woolf) one can only speak of a highly reduced interaction. When explaining how the reader understands that a book has been written by a person possibly with a communicative intent, IT can make use of narrative competency or the like.

When it comes to the error detection in the reader's mindreading of the real author, a comparison of the reader's final interpretation of the author's beliefs, intentions and emotions expressed in the text with the interpretations offered by other readers' (including critics and scholars), as well as with relevant authorial statements made outside the text in question, may help the reader evaluate the results of her mindreading. In particular, the author's statement might be privileged and function as a verifier or falsifier of the reader's interpretation. (*Nota bene*, in literary studies it is now most often assumed that there is no one correct interpretation of any text, interpretations being a result of interaction between the reader's mind and the text in a given cultural context, which means that there is no point of reference to verify or falsify the reader's attribution of mental states to the real author).

6. Conclusion: Possible effect of readers' mindreading experience on their competence in social understanding

In the light of TT, ST and IT, realist narratives, with their extensive and life-like presentation of characters' behaviour, offer the most life-like opportunity for mindreading practice; also the possibilities of error detection seem in this convention considerable. Ideal for mindreading practice (at least within ST and TT accounts) is a narrative that post factum reveals characters' mental states so that the reader's mental ascriptions can be verified/falsified (rather than merely dis/confirmed indirectly through verification/falsification of action predictions), provided that the author has good knowledge of human minds since this is fiction based on the author's knowledge, not authentic empirical material, that functions as a potential verifier/falsifier.

The wish to check whether one's guesses concerning the characters' experience and therefore their next actions are correct, i.e. consistent with the development of the story, might be the reader's primary motivation as regards literary mindreading. Compatibility or incompatibility between the reader's ascriptions and predictions and the text might constitute important feedback on the reader's mindreading. In particular, compatibility might be interpreted by the reader as a confirmation of her mindreading work and, therefore, proof that her mindreading faculties/resources are efficient. Incompatibility on the other hand, might be taken as a signal that her mindreading is faulty and needs to be modified to ensure better performance. Alternatively the reader might conclude that the author's mindreading system is flawed and this flaw is manifest in the implausible mental life and/or behaviour of

his characters. The reader might also take into consideration the possibility that both her and the author's mindreading faculties/competences are fine, but the character's behaviour/mental experience is nonstandard and hence unpredictable, or the work violates the verisimilitude principle (i.e. the characters are not meant to be life-like). Thus, the reader can spot incompatibility – possibly a signal of an error – but may be unable to locate its source. If the error derives from the reader's or the author's FP, enactment imagination or narrative competence, finding its proper location without reference to extra-literary sources may be impossible. Thus incompatibility falsifies the reader's ascription, prediction or anticipation but since it does not indicate the source of the error, it cannot automatically result in a modification (correction) of the reader's mindreading competence.⁹ Even so the chance to detect an error in one's mindreading in contact with literature provides some opportunity for the improvement, not merely practice, of the reader's mindreading capacity, though one must not forget that, as mentioned above, the strength of the falsifier/verifier can be questioned as it does not originate in real life but in the author's imaginative experience. Confirmation/verification of one's mindreading ascriptions, predictions and anticipations seems less problematic but is not unproblematic, as it may be based on a common mistake in the reader's and the author's mindreading (one can imagine, for example, that both the reader and the author are racists and systematically attribute ill intentions to members of ethnic minorities). Such mistakes will not be detectable without an extra-textual reference.

All these problems notwithstanding, literary mindreading seems to offer more extensive possibilities of error detection than real-life experience because a mistaken mental ascription or action prediction can be identified as such when it is compared with (often available) discussion of the character's mental experience or subsequent action presented by the (reliable) narrator, missing from real life. Conversely, some aspects of the mindreading experience are missing in contact with fiction, for example the reader's ability to monitor the other's mindreading of her (i.e. the reader), as neither the fictional character nor the real author can ever attempt to read the reader's mind.

When analysing the phenomenon of mindreading in the light of three cognitive theories – TT, ST and IT – with reference to fictional minds and in an imaginary social context that goes beyond the standard area in which these theories are normally applied, we hope to have also indicated a new method of examining mechanisms which according to the three theories govern our mindreading and social cognition. In the present article we have studied how TT, ST and IT deal with the question of errors that the literary mindreader can commit and detect. The choice of fiction, representing various conventions, gives a new dimension to the question and creates a unique platform for comparing and contrasting cognitive theories.

⁹ If the error is apparent and derives from the artefact's anti-mimetic character (e.g. the work's grotesque convention or the character's eccentric behaviour), the reader may be able to realize this on the basis of her literary competence.

References

- Apperly, I. A., Carroll, D. J., Samson, D., Humphreys, G. W., Qureshi, A., & Moffitt, G. (2010). Why are There Limits on Theory of Mind Use? Evidence from Adults' Ability to Follow Instructions from an Ignorant Speaker. *Quarterly Journal of Experimental Psychology*, *63*(6), 1201–1217.
- Bickhard, M. H. (2009). The Interactivist Model. *Synthese*, *166*(3), 547–591.
- Bickhard, M. H. (2013). The Social Ontology of Persons. In J. I. M. Carpendale & U. Muller (Eds.), *Social Interaction and the Development of Knowledge* (pp. 111–132). New York: Psychology Press.
- Birch, S. A. J., & Bloom, P. (2007). The Curse of Knowledge in Reasoning About False Beliefs. *Psychological Science*, *18*(5), 382–386.
- De Jaegher, H., & Di Paolo, E. A. (2007). Participatory Sense-Making: An Enactive Approach to Social Cognition. *Phenomenology and the Cognitive Sciences*, *6*(4), 485–507.
- De Jaegher, H., Di Paolo, E. A., & Gallagher, S. (2010). Can Social Interaction Constitute Social Cognition? *Trends in Cognitive Sciences*, *14*(10), 441–447. DOI: 10.1016/j.tics.2010.06.009.
- Di Paolo, E. A., Cuffari, E. C., & De Jaegher, H. (2018). *Linguistic Bodies. The Continuity Between Life and Language*. Cambridge: MIT Press.
- Dutton, D. (2009). *The Art Instinct: Beauty, Pleasure and Human Evolution*. New York: Oxford University Press.
- Fiebich, A., Gallagher, S., & Hutto, D. (2017). Pluralism, Interaction and the Ontogeny of Social Cognition. In J. Kiverstein (Ed.), *Routledge Handbook of Philosophy of the Social Mind* (pp. 208–221). London: Routledge.
- Froese, T., & Gallagher, S. (2012). Getting Interaction theory (IT) Together Integrating Developmental, Phenomenological, Enactive, and Dynamical Approaches to Social Interaction. *Interaction Studies*, *13*(3), 436–468.
- Gallagher, S. (2001). The Practice of Mind: Theory, Simulation or Primary Interaction? *Journal of Consciousness Studies*, *8*(5–6), 83–108.
- Gallagher, S. (2008a). Direct Perception in the Intersubjective Context. *Consciousness and Cognition*, *17*(2), 535–543.
- Gallagher, S. (2008b). Intersubjectivity in Perception. *Continental Philosophy Review*, *41*(2), 163–178.
- Gallagher, S. (2012). Empathy, Simulation, and Narrative. *Science in Context*, *25*(9), 355–381. DOI:10.1017/S0269889712000117.
- Gallagher, S. (2012). In Defense of Phenomenological Approaches to Social Cognition: Interacting with the Critics. *Review of Philosophy and Psychology*, *3*(2), 187–212. DOI: 10.1007/s13164-011-0080-1.
- Gallagher, S. (2014). An Education in Narratives. *Educational Philosophy and Theory*, *46*(6), 600–609. DOI:10.1080/00131857.2013.779213.

- Gallagher, S., & Hutto, D. (2008). Primary interaction and narrative practice. In J. Zlatev, T. Racine, C. Sinha & E. Itkonen (Eds.), *The Shared Mind: Perspectives on Intersubjectivity* (pp. 17–38). Amsterdam: John Benjamins.
- Gallagher, S., & Varga, S. (2014). Social Constraints on the Direct Perception of Emotions and Intentions. *Topoi*, 33(1), 185–199. <https://doi.org/10.1007/s11245-013-9203-x>.
- Gallagher, S., & Zahavi, Z. (2008). *The Phenomenological Mind: An Introduction to Philosophy of Mind and Cognitive Science*. New York: Routledge.
- Gallese, V., & Goldman, A. (1998). Mirror Neurons and the Simulation Theory of Mind-reading. *Trends in Cognitive Sciences*, 2(12), 493–501. [https://doi.org/10.1016/s1364-6613\(98\)01262-5](https://doi.org/10.1016/s1364-6613(98)01262-5).
- Goldman, A. I. (2013). *Joint Ventures Mindreading. Mirroring and Embodied Cognition*. Oxford: Oxford University Press.
- Goldman, A. I. (2006). *Simulating Minds: The Philosophy, Psychology and Neuroscience of Mindreading*. New York: Oxford University Press.
- Goldman, A. I., & Sripada, C. S. (2005). Simulationist Models of Face-Based Emotion Recognition. *Cognition*, 94(3), 193–213.
- Gopnik, A., & Meltzoff, A. N. (1996). *Words, Thoughts, and Theories*. Cambridge: MIT Press.
- Gopnik, A., & Wellman, H. M. (1995). Why the Child’s Theory of Mind Really is a Theory. In M. Davies & T. Stone (eds.), *Folk Psychology* (pp. 232–258). Oxford: Blackwell.
- Gopnik, A., & Wellman, H. M. (2012). Reconstructing Constructivism: Causal Models, Bayesian Learning Mechanisms, and the Theory Theory. *Psychological Bulletin*, 138(6), 1085–1108.
- Greene, G. (1981). *Twenty-One Stories*. London: Penguin.
- Herman, D. (2011). Introduction In D. Herman (Ed.), *The Emergence of Mind: Representations of Consciousness in Narrative Discourse in English* (pp. 1–40). Lincoln: University of Nebraska Press.
- Herman, D. (2011). 1880-1945: Re-Minding Modernism. In D. Herman (Ed.), *The Emergence of Mind: Representations of Consciousness in Narrative Discourse in English* (pp. 243–272). Lincoln: University of Nebraska Press.
- Hutto, D. D. (2007a). Narrative and Understanding Persons. *Royal Institute of Philosophy Supplements*, 60(5), 1–16.
- Hutto, D. D. (2007b). Folk Psychology Without Theory or Simulation. In D. D. Hutto and M. Ratcliffe (Eds.), *Folk Psychology Re-assessed* (pp. 115–135). Dordrecht: Springer.
- Hutto, D. D. (2008). *Folk Psychological Narratives: The Sociocultural Basis of Understanding Reasons*. Cambridge, Mass., London: MIT Press.
- Hutto, D. D., & Myin, E. (2017). *Evolving Enactivism: Basic Minds Meet Content*. Cambridge, Mass.: MIT Press.
- Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on Theory of Mind Use in Adults. *Cognition*, 89(1), 25–41.

- Lin, S., Keysar, B., & Epley, N. (2010). Reflexively Mindblind: Using Theory of Mind to Interpret Behavior Requires Effortful Attention. *Journal of Experimental Social Psychology*, 46(3), 551–556.
- Metzinger, T. (2014). First-Order Embodiment, Second-Order Embodiment, Third-Order Embodiment: From Spatiotemporal Self-location to Minimal Selfhood. In R. Shapiro (Ed.), *The Routledge Handbook of Embodied Cognition* (pp. 272–286). New York: Routledge.
- Mitchell, P., Robinson, E. J., Isaacs, J. E., & Nye, R. M. (1996). Contamination in Reasoning about False Belief: An Instance of Realist Bias in Adults but Not Children. *Cognition*, 59(1), 1–21.
- Newen, A. (2015). Understanding Others: The Person Model Theory. In T. Metzinger & J. M. Windt (Eds.), *Open MIND* (pp. 1–28). Frankfurt am Main: MIND Group. <https://doi.org/10.15502/9783958570320>.
- Nichols, Sh., & Stich, P. S. (2003). *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*. Oxford: Clarendon Press.
- Oatley, K. (2016). Fiction: Simulation of Social Worlds. *Trends in Cognitive Sciences*, 20(8), 618–628. <http://dx.doi.org/10.1016/j.tics.2016.06.002>.
- O'Neill, P. (1994). *Fictions of Discourse: Reading Narrative Theory*. Toronto: University of Toronto Press.
- Palmer, A. (2004). *Fictional Minds*. Lincoln: University of Nebraska Press.
- Quadt, L. (2015). Multiplicity Needs Coherence – Towards a Unifying Framework for Social Understanding. In T. Metzinger & J. M. Windt (Eds.), *Open MIND* (pp. 1–18). Frankfurt am Main: MIND Group. <https://doi.org/10.15502/9783958571112>.
- Ratcliffe, M. (2006). 'Folk psychology' is not Folk Psychology. *Phenomenology and the Cognitive Sciences*, 5(1), 31–52. DOI: 10.1007/s11097-005-9010-y.
- Rembowska-Pluciennik, M. (2018). Enacting Embodied Events in Narrative Processing. *Tekstualia*, 1(4) 145–157.
- Rushdie, S. (1982). Yorick. *Encounter*. September, 3–8.
- Schönherr, J., & Westra, E. (2019). Beyond 'Interaction': How to Understand Social Effects on Social Cognition. *The British Journal for the Philosophy of Science*, 70(1), 27–52.
- Stich, S., Nichols, S. (1992). Folk Psychology: Simulation or Tacit Theory? *Mind and Language*, 7(1–2), 35–71.
- Teske, J. K., & Gut, A. (2021). The Reader's Mindreading of Realist, Modernist, and Postmodern Fiction: A Comparative Study. *Narrative*, 29(1), 47–70.
- Thompson, E. (2007). *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Cambridge, Mass.: Harvard University Press.
- Wellman, H. M. (2014). *Making Minds: How Theory of Mind Develops*. Oxford, New York: Oxford University Press.

Woolf, V. (1923). Mrs. Dalloway in Bond Street. *Dial*, 75(1), 20–27.

Zahavi, D. (2001). Beyond Empathy: Phenomenological Approaches to Intersubjectivity. *Journal of Consciousness Studies*, 8(5–7), 151–167.

Zunshine, L. (2006). *Why We Read Fiction: Theory of Mind and the Novel*. Columbus: Ohio State University.

Joanna Klara Teske is Associate Professor at the John Paul II Catholic University of Lublin. She is the author of *Philosophy in Fiction* (2008), *Contradictions in Art: The Case of Postmodern Fiction* (2016), and articles on contemporary British fiction and cognitive theory of art, as well as co-editor of *Novelistic Inquiries into the Mind* (2016) and *Explorations of Consciousness in Contemporary Fiction* (2017). She is also an author of fiction and a publisher.

Arkadiusz Gut, PhD in Philosophy and Psychology, is Professor of Cognitive Science at The Nicolaus Copernicus University in Toruń. His research is focused on mindreading, thought-language relationship, and cross-culture study, and he directs an international research project focused on the involvement of language, culture-specific factors, and folk intuitions in social knowledge. His work has been published in *Multilingua*, *Journal of Cross-Cultural and Interlanguage Communication*, *Behavioral and Brain Sciences*, *Theory and Psychology*, *Frontiers in Psychology*, *Synthese*, *Narrative*.