



DRADEK

Studies in Philosophy of Literature, Aesthetics,
and New Media Theories

Vol. VI Num. 2 2020

ISSN 2465-1060
[online]

Self-Deception in Literature

Edited by
Zeynep Talay Turner

powered by



UNIVERSITÀ DI PISA

Comitato Direttivo/Editorial Board:

Danilo Manca (Università di Pisa, editor in chief), Francesco Rossi (Università di Pisa),
Alberto L. Siani (Università di Pisa).

Comitato Scientifico/Scientific Board

Leonardo Amoroso (Università di Pisa), Christian Benne (University of Copenhagen),
Andrew Benjamin (Monash University, Melbourne), Fabio Camilletti (Warwick
University), Luca Crescenzi (Università di Trento), Paul Crowther (NUI Galway),
William Marx (Université Paris Ouest Nanterre), Alexander Nehamas (Princeton
University), Antonio Prete (Università di Siena), David Roochnik (Boston University),
Antonietta Sanna (Università di Pisa), Claus Zittel (Stuttgart Universität).

Comitato di redazione/Executive Committee:

Alessandra Aloisi (Oxford University), Daniele De Santis (Charles University of
Prague), Agnese Di Riccio (The New School for Social Research, New York), Fabio
Fossa (Politecnico di Milano), Beatrice Occhini (Università di Napoli "L'Orientale"),
Elena Romagnoli (Scuola Normale Superiore di Pisa), Marta Vero (Università di Pisa,
journal manager).

ODRADEK. Studies in Philosophy of Literature, Aesthetics, and New Media Theories.
ISSN 2465-1060 [online]

Edited by Università di Pisa



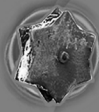
License Creative Commons

Odradek. Studies in Philosophy of Literature, Aesthetics and New Media Theories is
licensed under a Creative Commons attribution, non-commercial 4.0 International.

Further authorization out of this license terms may be available at <http://zetesisproject.com> or writing to: zetesis@unipi.it.

Layout editor: Stella Ammaturo, Marta Vero

Issue Editor: Zeynep Talay Turner



DRADEK

Studies in Philosophy of Literature, Aesthetics,
and New Media Theories

Vol. VI Num. 2 2020

ISSN 2465-1060
[online]

Self-Deception in Literature

Edited by
Zeynep Talay Turner

powered by



UNIVERSITÀ DI PISA

Self-deception and the Dynamics of Self-knowledge. Proust and Contemporary Philosophical Debate

Robert Pilat

Abstract

I examine the dynamic properties of self-knowledge that create an opportunity for self-deception. I argue that in self-deception, one mistakes the liberty concerning self-interpretation with the inability to deal with the dynamics of internal states. This inability affects three areas of self-governance: 1) managing belief change; 2) upholding the validity and force of past speech acts; 3) adjusting feelings to new situations and information. My analysis of self-deception presents self-knowledge as genuinely dynamic. To know oneself is to keep track of one's subjectivity and agency throughout the dynamic process of incorporating the old cognitive or emotional states into the new one. To deceive oneself is to take advantage of one's inability to govern this process in order to satisfy one's desires or reduce one's anxiety.

1. Legrandin's greeting and philosophy of self-deception

In a scene in Proust's *Swan's Way*, Mr. Legrandin, a family acquaintance, who is a snob, but also a person of some delicacy of feelings, arrives at a particular predicament.

Near the church, we met Legrandin, who was coming in the opposite direction escorting the same lady to her carriage. He passed close to us, did not break off his conversation with his neighbor, and from the corner of his blue eye gave us a little sign that was in some way interior to his eyelid and which, not involving the muscles of his face, could go perfectly unnoticed by the lady he was talking to; but seeking to compensate by intensity of feeling for the somewhat narrow field in which he had circumscribed its expression, in the azure corner assigned to us he set sparkling all the liveliness of a grace that exceeded playfulness, bordered on mischievousness; he overrefined the subtleties of amiability into winks of connivance, insinuations, the mysteries of complicity; and finally exalted his assurances of friendship into protestations of affection, into a declaration of love, illuminating for us alone, at that moment, with a secret languor invisible to the lady, a love-smitten eye in a face of ice¹.

Picking this scene as an example of self-deception is not original. Annette Barnes chose the

¹ Proust (2004), p. 128

same fragment before in her insightful study on self-deception². The choice of the same scene from Proust will help to integrate my point with the ongoing philosophical discussion on self-deception. Barnes is interested in the reasons Legrandin might have to think himself, not a snob. I am asking a related but not identical question, whether the gesture (the wink) made by Legrandin can be interpreted as a legitimate friendly greeting. A robust, intuitive feeling is to the effect that Legrandin commits self-deception. The question most commonly asked by philosophers discussing the phenomenon is this: How can rational subjects manage to deceive themselves? Several features of this act render it almost incomprehensible:

(1) One is at least partly aware of the very objects, actions, or facts one is so resolute not to know about.

(2) Discrepancies, possibly contradictions occur within the belief set when one disavows beliefs, which inferentially connect with the others;

(3) Social standards control most of our actions. Endowing an action with meaning, which violates the standards, turns the intended action into an entirely different action.

As far as the *first objection* is concerned, there has been an extensive philosophical discussion, which goes roughly between the intentional and the non-intentional approach to self-deception. The opposing positions are well represented by Donald Davidson (1986, 1994, 1998) and Mark Johnston

² Barnes (1997), p. 55

(1988) respectively³. On the intentional account, the self-deceiver holds two contradictory beliefs and virtually knows it. He is rationally obliged to assess the two claims and make an informed choice. Failing to do this, Davidson interprets as a reasoning failure. Since this kind of rational obligation makes the core of both rationality and self-knowledge, the self-deceiver, on Davidson's account, lacks both. Non-intentional interpretations make self-deception a matter of psychological mechanisms (probably with some discernible neural basis) blocking the right kind of information about oneself or forming a bias concerning some beliefs. Those theories usually remain silent about normative aspects of deceiving oneself (violating rationality, moral consequences). It is focused on the psychological plausibility of holding contradictory beliefs or withholding judgment despite overwhelming evidence.

Some philosophers search beyond the split. Carla Bagnoli (2007), for instance, focuses on the agency instead of mental operations. She suggests that self-deception amounts to disavowing the authorship of action (also the ownership of experiences and desires), thus failing to take responsibility for one's behavior. On such interpretations, self-knowledge is not, or not primarily, a cognitive activity. It is more like a *resolution* or *decision* to avow or disavow one's action.

The account I am presenting in this article takes something from all sides of the debate. On the one hand, I treat self-deception as anxiety-reducing like

³ The representative texts are collected in Rorty, ed. (1988).

Barnes, who subscribes to non-intentional view⁴. On the other hand, I claim that the reduction is accomplished by an effort of reflection, in an act which may be called prior rationalization. The self-deceivers prepare their systems of beliefs to deal with new information selectively.

As far as the *second objection* is concerned, Annette Barnes (1997) offers a very helpful insight. One can remove the discomfort caused by a belief by accepting many other beliefs, which contradict the former and reduce the pressure to maintain an undesired belief. However, the beliefs that make it possible to deceive oneself are available to consciousness only in an inferential way, i.e., as derived from other beliefs. A self-deceiver ascribes them to himself only *post factum*, after the discomfort and the resulting anxiety have occurred. One may compare it with someone entering an office where they have been working for years and seeing a big hole in the floor in the middle of the room. It would certainly be surprising, but not because the person previously believed there was no hole in the floor. That belief can be attributed to her only by inference. There is no independent testimony to the effect that she had it before. The view of a hole in the floor does not crash on the contradicting belief that there is no hole in the floor. Instead it shakes the whole set of stabilizing beliefs, which roughly define what is possible and what is impossible in the world. Although hard to itemize, this set of beliefs is not chaotic or random.

⁴ Barnes (1997), p. 33

Although the biasing (...) works non-intentionally, it has as its purpose anxiety reduction, and that purpose ensures that the biasing which occurs is neither random or haphazard. Furthermore, anxious desire does not automatically trigger biasing and self-deception. What other beliefs and desires, what other dispositions are operating in the agent, will affect when such biasing occurs⁵.

It is impossible to predict what beliefs will become the cause of the biased pursuit of the desired belief. Only when bias appears, a search starts for anxiety-reducing beliefs. On this account, the self-deception is a positive phenomenon and not only a lack of a cognitive state or a failure in reasoning. On the other hand, however, it is an irrational behavior as it ascribes the subject beliefs, which cannot be traced to their origin. In other words, they are entirely *ad hoc*. In section 2, I am offering an interpretation of the nature of this arbitrary move in probabilistic terms. It is an act of prior rationalization, which changes the distribution of probabilities in the set of data.

The *third objection* brings me close to the passage from Proust and to my main point in this paper. A greeting is a complex social activity that involves both external (social) interpretation and self-interpretation. Proust's ironic narrative leaves no doubt that Legrandin fails to convince the others about the plausibility of his non-standard greeting, but somehow he manages to convince himself. How is it possible?

⁵ Barnes (1997), p. 97.

Let W stand for Legrandin's gesture (the wink), and G stands for the success in presenting the gesture as a greeting. There are four possibilities:

(1) W and G	(2) $\neg W$ and G
(3) W and $\neg G$	(4) $\neg W$ and $\neg G$

All four possibilities should be taken into account by the rational subject in self-assessment, i.e., answering two questions: What do I do? What do I make it look like? All the probabilities – even the trivial (4) and somewhat exotic case (2) when the greeting is recognized as such no matter what Legrandin does – are constituent to the general likeness of Legrandin's success in social exchange. However, Legrandin gets to ignore inconvenient possibilities. Instead of aggregating the probabilities of variants 1 to 4, he focuses exclusively on the variant (1). By doing so, he violates the condition of rationality while preserving otherwise reasonable disposition.

Could he have done otherwise? Are human beings, in general, competent enough to be attentive to all relevant probabilities and to calculate the overall result? There is a body of evidence to the effect that they are not⁶. I stipulate that this incompetence is a part of a more general *human inability to deal with the dynamics of self-knowledge*. The difference between the common deficiency in probabilistic thinking and self-deception is that self-deceiver does not want to remedy this cognitive inadequacy.

6 Amos Tversky, Daniel Kahnemann, Paul Slovic, Sarah Lichtenstein, and their co-workers (1982) achieved rather conclusive results to this effect.

The last remark brings my approach closer to the intentional interpretation of self-deception. What I take from the non-intentional approach is the assumption about the role of temporal division in managing one's beliefs. Time is, indeed, an essential partition in our inner life. If all relevant information were available at the same time, self-deception would be unstable beyond usefulness – always on the verge of collapse, always in danger of disclosure. Its relative robustness, which we can attest daily, must have some other source. I stipulate that the self-deceiver transforms vice into a virtue by treating his *inability* to deal with the dynamics of self-knowledge as *liberty* in self-interpretation. In the remaining of this article, I will analyze three dynamic processes, which give both opportunity and means for self-deception: 1) managing belief change; 2) upholding the force of past speech acts; 3) adjusting feelings to new information.

2. Self-deception and the dynamics of beliefs

In this section, I am taking a closer look at belief change, and I am arguing that in many cases, like Legrandin's one, self-deception amounts to a defective *revision* of beliefs. Belief revision consists of adding new beliefs to one's belief set and retracting from this set all beliefs which are inconsistent with the new belief. The operation of selecting the beliefs to retract is far from trivial as numerous works in the

field clearly show. Self-deceivers fail to retract what they should or retract what they should not. However, such failures are not typical for self-deception; they also happen when we deliberately focus on defending a theory, or when we fail to learn from experience, or when we deliberately speak nonsense.

Self-deceivers cannot give reasons for inaccurate belief revision apart from saying: I wished it to be so, which renders them irrational. However, it is relatively easy to avoid the charge of irrationality if one refuses to revise their beliefs by disqualifying the oncoming information as irrelevant. To see how it might work, we need to step deeper into the logical and cognitive structure of belief revision.

The classical account of belief revision has been provided in the eighties by Carlos E. Alchourron, Peter Gärdenfors, and David Makinson (henceforth AGM), followed by extensive discussion and extensions. I want to show how to interpret self-deception in the domain of beliefs by resorting to this formal apparatus. The AGM model sets conditions for belief changes that preserve the logical structure of a set of beliefs while admitting new information. The sets of beliefs have to be closed on the consequence. If new information is inconsistent with the old belief set, something has to be retracted (Legrandin has more than enough social knowledge to prevent him from thinking about his strange gesture as a legitimate greeting). This retraction must not be massive as it would undermine the person's rationality (Legrandin should not retract all his previous beliefs about greeting friends in public).

Rational subjects perform belief revision according to a specified ratio: addition-retraction. “Additions and retractions trade-off against each other”⁷. Belief change should be performed orderly given a specific revision function for a belief set *S*. This function has to satisfy formal AGM postulates. If this is the case, the sentences in *S* can be ordered according to their doxastic preferability. The criteria would be strength, entrenchment, comparative retractability⁸. Now, the crucial thing is to find the appropriate revision function.

Gärdenfors (1990) examines a basic criterion of retractability based on probability measures. He starts from formulating a rule: “If a belief state *K* is revised by a sentence *A*, then all sentences in *K* that are irrelevant to the validity of *A* should be retained in the revised state of belief”⁹. Now a suitable definition of relevance is needed. Gärdenfors starts from classic Maynard Keynes’ definition:

A is relevant to C iff $P(C/A) \neq P(C)$

A is irrelevant to C iff $P(C/A) = P(C)$

This definition is then tested against logical (syntactic) criteria of relevance relation (**R** stands for relevance, **IR** for irrelevance)

(1) (R0) If $\vdash A \leftrightarrow B$, then $A \mathbf{R} C$ iff $B \mathbf{R} C$.

(2) (R1) $A \mathbf{R} C$ if not $A \mathbf{IR} C$

(3) (R2) $A \mathbf{R} C$ if not $\neg A \mathbf{IR} C$

(4) (R3) $(A \vee \neg A) \mathbf{IR} C$

(5) (R4) If *C* is contingent, then $C \mathbf{R} C$

7 Schulte (1999), p. 144-155

8. Rott (2008), p. 125

9 Gärdenfors (1990), p. 349

(6) (R5) If $A \mathbf{R} C$ and not $\vdash \neg(A \& B)$, then $(A \& B) \mathbf{R} C$.

(7) (R6) If $A \mathbf{R} C$, $B \mathbf{R} C$, and not $\vdash \neg(A \& B)$, then $(A \& B) \mathbf{R} C$.

(8) (R7) If $A \mathbf{I} \mathbf{R} C$ and $B \mathbf{I} \mathbf{R} C$, then $(A \& B) \mathbf{I} \mathbf{R} C$.

The original Keynes' definition meets criteria R0 to R4, but these alone render too few beliefs relevant, hence we should add criteria. R5 to R8. Now Gärdenfors tries to find out what definition of relevance may serve as minimal revision function for a set of beliefs when the holder is confronted with a new piece of information. He arrives at what he calls his final definition of relevance. In his text, it is marked D4 and says the following:

(a) $A \mathbf{I} \mathbf{R} C$ iff $P(A) = 0$, or for all B such that $P_{-B}^-(C) = P_{-B}^-(C/B)$ and $P_{-A}^-(A \& B) \neq 0$, it also holds that $P_{-A}^-(C/A \& B) = P_{-A}^-(C)$.

(b) $A \mathbf{R} C$ iff not $A \mathbf{I} \mathbf{R} C$

Symbols P_{-A}^- and P_{-B}^- in this definition denote *contracting functions* that yield a probability value between 0 and 1 for such an A , which before the change was assigned probability value 1. The definition of relevance (irrelevance) can be formulated as a set of conditions of retracting belief A if C is retracted. The formalism reads as follows: A is irrelevant in relation to C if and only if there is no difference between two probabilities: 1) unconditional probability of C contracted upon acceptance of B ; 2) the conditional probability of C contracted upon acceptance of B given A . If A is so irrelevant then if C is retracted from belief set, A should stay albeit contracted upon new information.

Gärdenfors searches for rules of contraction, but he concludes that

the postulates do not determine a unique contraction, but they only introduce rationality constraints on such functions (...) rationality constraints are not enough to determine the unique contraction function (...), but pragmatic factors must be added in order to single out the actual contraction¹⁰.

Gärdenfors suggests that the pragmatic factors come from *epistemic entrenchment*.

The crucial problem in belief revision is the ordering of entrenchment values of different beliefs. This ordering cannot be described formally. For instance, Hans Rott defines the cognitive preference of one belief over another belief by reference to the *belief basis*¹¹. The basis is understood here as a set of independent pieces of information. One may define a cognitive priority as a relation between subsets of a belief basis. The author considers two such relations, which constitute negative and positive notions of entrenchment marked $\blacktriangleleft+$ and $\blacktriangleleft-$. The positive notion takes the weakest element of corresponding sets as a criterion and can be defined as follows ($<$ stands for preference relation)

$G \blacktriangleleft+ G'$ iff G is not empty and for every χ in G' there is a ξ in G such that $\xi < \chi$

The negative notion takes the strongest element in certain complements $G - G'$ and $G' - G$ (remainder

¹⁰ Gärdenfors (1990), p. 359

¹¹ Rott (1995)

sets consisting of the pieces of information one may retain when forced to give up some beliefs) and can be defined as follows:

$G \leftarrow G'$ iff $G \neq G'$ and for every χ in $G - G'$ there is a ξ in $G' - G$ such that $\chi < \xi$

The author shows that the negative notion does better in explaining how the priorities in belief base influence the cognitive preference in belief sets under revision.

Self-deceiver fails to revise his belief set properly due to particular *prior rationalization*. The question is what this rationalization does. I stipulate that it must have something to do with assigning probabilities and with prioritizing the belief basis. The self-deceiver distorts the Bayesian distribution of probabilities. It is possible because people can produce concepts *ad hoc*, and adding new concepts changes the distribution of probability. It is like adding one more possible action to the set of activities from which we have to choose. The distribution of probabilities over the whole set changes, and all choices are affected.

The problem is not that the self-deceiver does not calculate probabilities or does it poorly - most people are that way. It has to do with refusing to see the probabilistic structure of information. The actual calculation is not necessary to admit the presence of a network of probabilities. Taking them into account may take a form of guessing, caution, hesitation, or double-checking the status of the situation. None of these appear in the behavior of Legrandin or similar self-deceivers.

Why is it so easy not to see the information structure even though all constituent information is readily available? The most plausible explanation seems to be the following: In order to deal with distributions of probabilities, we need to be active users of the *concept* of probability. However, we hardly resort to this concept in daily life as it is not part of our natural abilities (there was no science of probability before the XVII century). Other concepts easily overpower the concept of probability. For instance, in Legrandin's mind, the concept of probability is pushed out by that of intentional causality. In the case of pure physical action, we readily accept that we may fail to do what we intended, but in the case of social gesture, we usually think that we do what we mean. In other words, we assume that our intentionality determines the *content* of our acts. Since this causality is perceived as very strong, it is hard to change the causal reasoning into a probabilistic one (and consequently to allow for the possibility of failure). The probability distribution (in this case, it is a probability distribution of social perceptions of Legrandin's gesture) is absent or distorted. Indeed, as we know from Kahneman's and Tversky's findings, the heuristics based on causality is a powerful bias making us blind to the Bayesian structure of uncertainty. That is why Legrandin comes to terms with his unusual gesture all too easily.

3. Self-deception and the dynamics of speech acts

The dynamic nature of beliefs and probabilities is not the only dynamic factor in self-deception. As I said in section 1, two more dynamic processes have to be considered: 1) preserving the validity of the past speech acts; 2) recognizing and updating feelings. In this section, I discuss the illocutions.

In Georges Bizet's opera *Pearl Fishermen*, two friends, Nadir and Zurga, are in love with Leila but decide not to pursue their passion in the name of their friendship. They make each other a promise to this effect, but Nadir does not keep his word. Now, there are three possibilities:

1) Nadir deceived Zurga when he made his promise.

2) Nadir deceived himself as to the state of his feelings, and only later, his self-illusion broke. He realizes that he should not have made the promise in the first place.

3) Nadir made an honest promise, but he was in the wrong in thinking that he would be in control of his emotions later on.

Zurga accepts the most charitable third interpretation of his friend's conduct. Consequently, when the truth about the romance comes out, he rescues his friend (sacrificing his life) from the vengeance of the community, which condemns the romance as blasphemous (Leila is a priestess). Zurga would probably not have made such a sacrifice if he had chosen the first interpretation. He would also

be less inclined to help if he had chosen the second one. However, Zurga did not have enough evidence for either of the interpretations. His charitable choice was arbitrary - based on friendship alone.

Did the friends know what they were doing when they made their vows? In an unparalleled duet of both characters (Act 1: *Au fond du temple*), they recall their first encounter with Leila and the tension that grew between them, and how they decided to reject all of this, to remain friends forever. Now we know. The musical language does not lie: the promises made by both of them were sincere but equally genuine is what we later learn from Nadir's aria *A cette voix ... Je crois entendre encore*, namely, that immediately after making a promise, he did seek contact with Leila. So what was the nature of Nadir's deed? Was he in a state of self-deception while promising? If so, was it a valid promise? If the promise was not valid, Nadir could not possibly break it. Is every promise void if based on self-deception?

A promise is an act that we believe should spread over the thoughts, actions, and intentions of the person who promises. When this does not happen, we can consider the promise void, or we can say that the promise was broken. Interestingly, Nadir himself is inclined to the least charitable self-interpretation, accusing himself of betraying his friend by making a false promise. In other words, he projects his guilt backward to the moment of promising. However, the duet of Nadir and Zurga mentioned above contradicts his self-accusation. For all we know, the promise was indeed sincere; they both thought at that

moment that their friendship had enough strength to ensure the pact's durability. At the time of the pledge, nothing had yet been decided. Erroneous self-knowledge and self-deception are *dynamic phenomena* that do not arise immediately but through a gradual strengthening. Only after some time, it turns out what we did or meant. However, in opera, not unlike in real life, critical decisions have to be made before the final truth is unveiled. So everything depends on interpretation: Zurga must decide whether he is dealing with Nadir's lie or Nadir's self-deception or Nadir's error. He must take a risk. The decision here is based on probabilistic premises. The possibility of a fundamental moral defect in Nadir is not entirely excluded until the very end of the opera. The viewers have to decide for themselves. This openness is what makes Bizet's work so compelling.

The search for criteria and cognitive foundations of self-deception must go beyond looking into the mind of the belief holder (this strategy is typical for philosophical-analytic studies of self-deception). Our beliefs, as well as illocutionary acts and feelings, are always unfinished, half-constituted, elusive. Both the alleged self-deceiver and the interpreter have to *decide* what they are dealing with. In making these decisions, they must remember what motivates the interpreter. Zurga had chosen a charitable interpretation of Nadir's behavior not because he had good reasons to do this, but because Nadir was his friend. Nadir, on the other hand, was driven by remorse, so he was more inclined to see himself as a self-deceiver, dishonest both to himself and his friend.

4. Self-deception and the dynamics of feelings

Let us consider the third dynamics of internal states, which notoriously provide opportunities for self-deception: updating one's feelings upon receiving some vital information about one's properties or situation. The growth of self-knowledge can lead to unexpected modifications of feelings. Some feelings disappear in the full light of consciousness, and others are considerably changed: strengthened, fixed, assigned new function, elevated or lowered in their relative status. These changes are always somehow unpredictable. Persons who are skilled in introspection and are sensitive to signals from others can deal with the adjustment of their feelings with some fluency, but even they do not fully control the process. Even the most well-established feelings become uncertain and fleeting under the influence of reflection. Their dynamics are particularly challenging to understand:

The aporia shown here is even more profound and more acute. Martha Nussbaum shows it insightfully in her essay *Love's Knowledge*. She analyses a fragment from *Albertine Gone*, in which Marcel learns that his lover Albertine abandoned him and left.

Yet, a moment ago, before Françoise came into the room, I had believed that I no longer loved Albertine, I had believed that I was leaving nothing out of the account, like a rigorous analyst; I had believed that I knew the state of my own heart. But our intelligence,

however lucid, cannot perceive the elements that compose it and remain unsuspected so long as, from the volatile state in which they generally exist, a phenomenon capable of isolating them has not subjected them to the first stages of solidification. I had been mistaken in thinking that I could see clearly into my own heart. But this knowledge, which the shrewdest perceptions of the mind would not have given me, had now been brought to me, hard, glittering, strange, like a crystallized salt, by the abrupt reaction of pain¹².

A sudden realization of a feeling has consequences for the perception of the past feelings of the same person. The feeling presents itself now as persisting already in the past. Marcel just a second ago has not avowed any feelings for Albertine in his past; now, he realizes he has had those feelings all along. Let us use an analogy: The claim to retrospective meaning is characteristic for time awareness as such, especially in the perception of rhythmic phenomena. One may be unaware of a ticking clock. However, when suddenly the ticking makes it to the consciousness, one also realizes that it has been audible all along (the example was given by the Polish phenomenologist Roman Ingarden). It seems that the phenomenological structure of this experience involves three stages: 1) no memory; 2) quasi-remembrance of the ticking of the clock; 3) conviction of the ticking of the clock for some time. Quasi-remembrance seems to be of critical

¹² Proust (1981), III, p. 426

importance here. It is similar to judgment because it gets accomplished by inferring the past ticking of the clock from the current ticking. However, a person who has such a *retroactive impression* certainly does not feel that she is inferring. Instead, she hears a particular acoustic line stretching backward – something that originates in the present experience and stretches along some virtual acoustic line into the past. However, one has to use analogies with caution. The difference between the clock example and the feeling example is essential: the remembrance of feelings or lack thereof makes a significant impact on whom the person considers herself to be.

The analogy with sound perception tells us that feelings in the past should not be seen in the light of the current impressions. Firstly, it would be entirely unverifiable to move from an immediate impression to a statement of a past feeling. Secondly, the feelings we experience now are usually evaluated by considering their relation to past feelings. However, such evaluation is only credible if the past feelings can be judged on independent evidence, and not based on inference from the present impressions.

Proust insightfully describes Marcel's situation as a search for the truth about his feelings. The sudden realization of a feeling throws him into the different present and by implication into different past. Marcel has no resources to settle between the two images of his emotional history. His past gets doubled, which constitutes a severe aporia in the domain of self-knowledge.

Martha Nussbaum presents three solutions of the aporia. She takes the first two from Proust himself. They contain an essential conceptual distinction that separates the current experience (the awareness of the feeling) from the object of the experience (the feeling itself). They also suggest that the current experience is a reliable source of knowledge about the feeling which endures in time. The third solution is also based on a literary text - a short story by contemporary American writer Anne Beatie *Learning to Fall*. To the solutions analyzed by Nussbaum, I am adding two more proposals: the first one inspired by Max Scheler, the other by Harry Frankfurt.

First solution. The intellectual analysis based on self-observation and leading to the development of “the science of one’s soul” is decisive in establishing the truth about one’s feelings. The experience of sudden suffering makes Marcel realize that he must have loved Albertine after all. Pain has a kind of intentionality; it points to its cause. However, this feature is difficult to understand. The anguish could be caused by anything: offended ambition, regret at the hardships, or moral indignation. That is why Marcel wants to know more: does suffering tell anything of consequence? If so, does it tell the truth? Looking for an answer, he uses an analogy: the knowledge of a feeling is like a crystal of salt, created by a chemical reaction. The suffering catalyzes self-knowledge.

Nussbaum attempts to give this analogy a philosophical meaning. She invokes the old stoic teaching about cataleptic impressions. Stoics claimed that there are such forms of experience that infallibly,

without additional verification, inform about their objects. Its credibility is supposed to come from three properties: uncontrolled force, the passivity of the subject, and suddenness. Can they be a foundation of the Proustian “science of one’s soul”? There are five doubts I can think of:

1. If suffering were a paradigmatic case of cataleptic experience, it would probably work much better for negative feelings. The self-knowledge of purely positive feelings, like love, joy or happiness, would be difficult to imagine.
2. Suffering is not specific enough. It may result from a large class of possible causes. It cannot, therefore, be the basis for a verifiable judgment about one’s feelings.
3. Experiences do not say much without context. In the case of Marcel and Albertine, Proust meticulously describes the context: the turbulent history of the relationship, the characters of both lovers, the norms, and social conventions. However, if the context is necessary, then the knowledge of one’s feelings is, to a great extent, based on the collection of external information about oneself and not on the internal perception of one’s feelings.
4. There are many phenomena closely related to love (such as tenderness or care) that do not seem to evoke any cataleptic sensations. Catalepsy would grasp only some salient aspects of such feeling, which may not be

the most important or central. In other words, sometimes we recognize our feelings by marginal characteristics, which can be very confusing.

5. The very search for cataleptic sensations is motivated by an intellectual aspiration which is born far away from emotional life, namely in scientific pursuit.

Is the claim to emotional self-knowledge realistic at all? If it is not, then the danger of self-deception is in the very center of the famous “know thyself”.

Second solution. Nussbaum points out that Proust finally makes a compromise between an intellectual approach and a non-discursive reference to sensations. In *Time regained* Proust unfolds reflections on this subject in connection with his thoughts on the work of art. By exploring ourselves, we discover specific patterns, which tell us the truth about ourselves. The shape of the work and the shape of one’s own life emerge at the same time. Feelings have a special place in this dynamics as they are creating continuity beneath all those patterns. Feelings add their flow to this dynamics. Proust speaks of the end of a feeling as a necessary condition for another similar feeling, not unlike Adina in Gaetano Donizetti’s *L’elisir d’amore*, who urges Nemorino (*Chiedi all’ aura lusingheira*) to take her example and let the unhappy love be driven out by a new love. The invariants in such dynamic processes are supposed to be the subjects themselves – they are supposed to sustain in these transformations, loyal to themselves, not to any particular feelings. This solution, however, sounds like changing the

game: we are not after self-knowledge anymore; we are busy protecting our identity.

Third solution. Nussbaum believes that any project of creating a science of the soul is hopeless. Experience and intellect are in a constant struggle to dominate self-interpretation. None of them can break free from the control of another. Nussbaum seeks a solution that is free from the “science of the soul” and replaces Proustian (Cartesian) self-knowledge with *self-knowledge through engagement*. A short story by Ann Beattie *Learning to Fall* inspires her interpretation. Knowledge about love is not pure discovery. On Nussbaum’s account, actions, and experiences intertwine. We do not know precisely where the performative act (throwing ourselves into a situation) ends, and where the knowledge about this situation begins. This fragile boundary must be protected especially by an attentive and sensitive distance, sometimes irony, sometimes humor; in any case, it is not merely a matter of what one’s experience says¹³.

Fourth solution: According to Max Scheler, feelings are intentional objects characterized by insatiability. They always appear accompanied by the awareness of unfulfillment, of something missing in their intentional structure. The intentional object or content of a feeling is never entirely given. It should instead be thought of as an ideal object. As far as the human feelings for each other are concerned, such objects provide a model of an ideal emotional community. In other words, in any living experience

13 Nussbaum (1990), p. 504-505

of a feeling, there is a sense of an ideal community of feelings. Empathy is, therefore, a component of every feeling as well as it is a spiritual act aimed at the ideal emotional objects¹⁴.

Adding an empathy condition makes feeling depend to no small extent on social conventions and training. Before experiencing the pain of Albertine's departure, Marcel had a firm conviction about the content of his feeling for Albertine; namely, he thought it was toxic and everything but love. This conviction was mistaken, as the violent experience of Albertine's departure revealed. This experience, however, did not improve his cognitive situation significantly: now, he knew that he loved her, but he is less than ever sure about the *content* of that feeling. It appeared to him as a mystery, an incomprehensible, fearful state – real to be sure, but of unknown sense and implications. So Marcel could not think of his discovery as bringing about any self-knowledge; at most, it was a realization that self-knowledge was absent before.

Fifth solution: Harry Frankfurt in *The Dear Self* wonders whether the expression “self-love” has a literal and not just figurative meaning. He makes a surprising observation: Self-love fulfills very well the conditions of love as such:

1. It focuses on the person who is loved.
2. It desires her good and her sheer existence.
3. It is selfless as there is no interest here beside the benefit of the loved person
(Frankfurt is a bit sophistic here).

¹⁴ Scheler (1923), pp. 228-232

The difference between love and self-love exists not in the characteristics of the feeling, but in the fact that it is difficult to point out the specific act directed at oneself that would realize self-love. For example, when I want a good for myself, this desire is accomplished exclusively by my desire for this good, and not by a separate act of desiring this good for myself. A paradoxical situation therefore arises. Based on Frankfurt's criteria, we can describe the content of a positive relationship with ourselves, but we are not able to demonstrate that this is indeed a relationship with ourselves. The active end of this relationship is missing.

Frankfurt has some answer to this difficulty: No act directed toward oneself is necessary. We only need a certain quality of the relation, namely *wholeheartedness*, the sincere, undivided, internally undisturbed act directed at the good that one wants for themselves. If the subject directs himself entirely towards this good, and the desire itself fulfills the general conditions of love, then it is also an act of self-love. The subject becomes the complete object of his procedures when he is the wholehearted subject of those procedures. In other words, the highest good for the loved one is to be a perfect loving one. In other words, good things bind us to ourselves as long as our commitment to them is wholehearted.

This solution has its advantages, but it is void of any normative character. After all, full commitment alone can also occur in the case of apparent goods, of which we will later feel ashamed. It is too high a

price to pay for an elegant explanation of emotional self-relation and emotional self-knowledge.

As we saw, all five solutions to the problem of emotional self-knowledge have some shortcomings. The passage from the experience of a feeling to the emotional self-knowledge continues to be elusive. It leaves vast space for such self-interpretation in which everything seems possible for the subjects. The richness of possibilities makes the impression of infinite freedom concerning one's internal states, thus encouraging the person to venture arbitrary self-interpretations. Alas, this freedom is not the result of cognitive power, but cognitive incapability.

5. The ever open space for self-deception

In self-deception, we show incompetence in governing the threefold dynamics: cognitive (belief change), social (sustainability of illocutions), and emotional (updating feelings). The incompetence can lead to a stable state of self-delusion concerning one's beliefs, commitments, and emotions.

The first incompetence is about probabilities. Here we do not commit self-deception by doing something, but rather by failing to do something, namely, we do not improve our poor natural competence in probabilistic thinking - we stick to our inherent cognitive weaknesses. It might explain why self-deception is sometimes so easy. It is different in this respect from lying, which is a hard job (keeping

track of lies to avoid inconsistencies). Humans are natural self-deceivers – we do not notice that, because slight deviations from the truth about ourselves are so frequent. We take notice only when they become grave.

The second incompetence concerns illocutionary acts. Social life require that the illocutionary force of any previous speech act (promises, thanks, declarations, confessions) is maintained. If the speakers fail to secure that, the others are forced to choose if they have to do with a lie, self-deception of just a psychological instability. The interpreters have to take a risk. The risk also concerns the self-interpretation of the actor. We interpret ourselves in light of some external, conventional criteria, which leaves space for self-deception, especially about such acts as promises, commitments, declarations, thanks, pleas.

The third incompetence concerns the nature of emotional self-knowledge. Our emotions are known to us in retrospection, but we often mistake them for actual states. When we perceive ourselves, we always look backward into a specific pattern emerging from the past, so we learn from Proust. However, precisely how far back should we look? It is a delicate matter. If one reaches too deep into the pattern, one ends up clinched to a rigid, dogmatic view of one's emotional life. Self-perception becomes a theory about oneself. The content of feelings is no more constituted in the first person perspective. It might even become so distant that it loses personal character entirely. If, however, the gaze is too shallow, one finds himself

at the mercy of fleeting impressions. Both cases constitute a ground for self-deception. The former does so by opposing any new information about oneself. The latter does so by voiding this information (depriving it of weight or relevance).

Thoughts, beliefs, feelings, perceptions are dynamic entities. The inability to govern them creates the desire, and the opportunity for self-deception, and sometimes even means to accomplish it.

Bibliography

- Bagnoli, C. (2007): "The Authority of Reflection", «Theoria», vol. 22, No.1, 2007, pp. 43-52.
- Barnes, A. (1997): *Seeing Through Self-Deception*, Cambridge: Cambridge University Press.
- Davidson, D. (1994): *Knowing One's Own Mind*, in: *Self-Knowledge*, edited by Q. Cassam, New York: Oxford University Press, pp. 43-64.
- Davidson, D. (1986): *Deception and Division*, in J. Elster (ed.), *The Multiple Self*, Cambridge: Cambridge University Press.
- Davidson, D. (1998): *Who is Fooled?* in J-P. Dupuy, *Self-Deception and Paradoxes of Rationality*. Stanford: CSLI Publication.
- Gärdenfors, P. (1990): "Belief Revision and Relevance" in «PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association», vol. 2, 1990, pp. 349-365.
- Johnston, M. (1988): *Self-Deception and the Nature of Mind*, in O. A. Rorty, B. P. McLaughlin (ed.), *Perspectives on Self-Deception*, Berkeley: University of California Press, pp. 63-91.
- Kahneman, D., Slovic P., Tversky A. (1982): eds., *Judgment under Uncertainty: Heuristics and Biases*, Cambridge: Cambridge University Press.
- Moran, R. (2001): *Authortity and Estrangement. An Essay on Self-Knowledge*, Princeton: Princeton University Press.

- Nussbaum, M. (1990): *Love's Knowledge. Essays on Philosophy and Literature*, New York: Oxford University Press.
- Proust, M. (1981): *Remembrance of Things Past*, trans. By C. K. Scott Moncrieff and A. Mayor, (revised by Terence Kilmartin), New York: University of Wisconsin Press.
- Proust, M. (2004): *Swann's Way*, trans. edited by L. Davis, London: Penguin Books.
- Rorty O. A., McLaughlin B. P. (1988), *Perspectives on Self-Deception*, Berkeley: University of California Press.
- Rott, H. (1995): "Just because". *Taking Belief Bases Seriously*, in S. Hanson, W. Rabinowicz, K. Segerberg (eds.) *Logic for a change. Essays Dedicated to Sten Lindström on the occasion of his 50th Birthday*, Uppsala: Uppsala Prints and Preprints in Philosophy.
- Rott, H. (2008): "A New Psychologism in Logic? Reflections from the Point of View of Belief Revision", «*Studia Logica*» vol. 88, p. 125.
- Scheler, M. (1923): *Wesen und Formen der Sympathie*, edited by M. S. Frings, Bern: Francke Verlag, 1973.
- Schulte, O. (1999): *Minimal Belief Change and Pareto Optimality*, in *Proceedings of the 12th Australian Joint Conference on Artificial Intelligence: Advanced Topics in Artificial Intelligence*, London: Springer Verlag.

