

# HUNGARIAN PHILOSOPHICAL REVIEW

VOL. 64. (2020/1)

---

The Journal of the Philosophical Committee  
of the Hungarian Academy of Sciences

Self, Narrativity, Emotions

Edited by Gergely Ambrus and Csaba Olay



# Contents

Preface	5
CSABA PLÉH: Narrative Identity in its Crises in Modern Literature	9
GÁBOR BOROS: First Phase of the Narrative Theory of Personal Identity: Wilhelm Dilthey, and Georg Misch	24
TIM THORNTON: Narrative Identity and Dementia	40
GERGELY AMBRUS: The Identity of Persons: Narrative Constitution or Psychological Continuity?	59
JUDIT SZALAI: The ‘Reasons of Love’ Debate in Analytic Philosophy: Reasons, Narratives, and Biology	77
DAVID WEBERMAN: What is an Existential Emotion?	88
LORE HÜHN: Com- <i>passion</i> : On the Foundations of Moral Philosophy for J. J. Rousseau and Arthur Schopenhauer	101
HYE YOUNG KIM: An Existentialist Analysis of Forgiveness and Gratitude	117
JAMES CARTLIDGE: Heidegger’s Philosophical Anthropology of Moods	126
PHILIPPE CABESTAN: Bad Faith <i>versus</i> Unconscious: a Credible Alternative?	141
CSABA OLAY: Alienation	149
PHILIPP HÖFELE: New Technologies and the “Heuristics of Fear”: The Meaning and Prehistory of an Emotion in Jonas, Heidegger and Hegel	166
Contributors	183
Summaries	185



## Preface

The present volume of the *Hungarian Philosophical Review* addresses current issues in the philosophy of self. The contributions may be divided into two larger groups. The first set of papers discusses themes that evolved in the frames of the narrative approach to self and personal identity. The second group is organized around the idea of existential emotions (e.g. *angst*, guilt, compassion, gratitude, forgiveness), the sort of emotions assumed to be fundamental for being an individual self (or a *Dasein*). The idea that both narrativity and existential emotions play an essential role in characterising the nature of human existence has become prominent in the last few decades.

The section *Narrativity and Self* begins with Csaba Pléh's paper which provides an overview of the origins of narrative theories of the self. Among the diverse sources Pléh lists psychological research on narrative memory, initiated by Jerome Bruner, philosophical theories suggesting a narrative construction of the self, promoted e.g. by Paul Ricœur and Daniel Dennett, modern novelists and literary theorists from Milan Kundera to David Lodge who proposed novel writing as a factor in the birth of the modern self, and also research on specifically autobiographical narratives that present the unfolding of the self in autobiographical story-telling practices. He concludes, agreeing with Bruner, that the search for explanatory principles underlying schemata by the experimentalists, the use of autobiographical narratives, and the cultivation of broken narrative patterns in modern novels can be seen as a modern way to present the traditional dualism of *Naturwissenschaften* and *Geisteswissenschaften* as a duality of a categorical and a narrative approach to the human mind.

Gábor Boros' paper addresses the particular role research on autobiography played in narrative theories of identity. Boros notes that contemporary narrativist theories of identity rarely mention Wilhelm Dilthey and Georg Misch, two German philosophers, active between the mid-19<sup>th</sup> and the mid-20<sup>th</sup> century. Dilthey and Misch, in Boros' opinion, were notable forerunners of these contemporary movements, whose views deserve to be integrated into the history of narrative identity movement. Boros also argues that the theories of Dilthey

and Misch are not only interesting from a historical point of view, but may also be seen as providing new ideas to the contemporary discourse on identity and narrativity.

Tim Thornton investigates another aspect of the narrative approach to self and personal identity, namely its possible applications to dealing with patients with Alzheimer disease and other types of dementia that lead to a diminished presence of selfhood and personal identity. Thornton starts his discussion with the longstanding view that personal identity depends on memory, and since dementia causes serious deterioration of memory functions, hence it undermines personal identity. He, then, draws attention to views of philosophers and health-care professionals who criticised this connection, relying on a narrative account of identity. These critics maintain that while the capacity to author a self-narrative is threatened by dementia, personal identity may nonetheless be saved if the relevant narrative can be *co-constructed* with others. Thornton explains the dangers of any such co-constructionist proposals, and also suggests an alternative, minimal account of what role narratives in dementia may play, making use of Wittgenstein's notion of secondary sense.

Gergely Ambrus addresses the philosophical debate between two strongly opposing approaches, the psychological continuity and the narrativist theories of the self and personal identity. In particular, he examines Marya Schechtman's narrative self-constitution view and contrasts it with Derek Parfit's neo-Humean psychological continuity theory. Ambrus sets out to defend Parfit against a major criticism of Schechtman which seeks to discredit Parfit's notion of quasi-memory (and quasi-belief, quasi-desire etc. as well). Parfit's psychological continuity view essentially depends on these q-notions, hence undermining them provides a ground for accepting narrativism. Despite defending it from Schechtman's attack, the author also argues that the psychological continuity view fails seriously, as it does not account for identification he takes to be a necessary condition of being the same person. The paper concludes by considering some possible explanations of identification, and by considering whether they support the narrativist or the psychological continuity view.

Judit Szalai's paper is a contribution to the "reasons of love" debate in analytic philosophy, and in laying out her own position she draws upon the role of narratives in making distinctions between different forms of love. She argues for the following tenets. The opposition between "reasons-based" and "no-reason" views does not constitute a genuine theoretical dilemma: we do not love persons for either abstract properties that several individuals can share, or for some elusive "ipseity". Further, she also stresses that descriptive and normative approaches concerning love, viz. why persons in fact love others and why they should love must be clearly distinguished. Third, distinguishing between different types of love is important, since reasons apply to these in different way. Lastly, she also shows how the interplay of different factors in loving persons,

such as personal properties of the beloved as reason-giving, the joint history, and bio-psychological factors are relevant in understanding romantic love.

The section *Self and Existential Emotions* begins with David Weberman's paper attempting to delineate what is the subset of emotions that might be qualified as existential. He takes as point of departure Heidegger's account of affectivity in *Being and Time*, while adjusting the terminology and developing the conception in directions Heidegger did not explore. The paper examines Heidegger's notion of *Befindlichkeit* as a description of two types of what we call emotions: moods and object-specific emotions. The importance of moods lies in that moods bring us up against the fact that *Dasein* is delivered over ("*ueberantwortet*") to being and consequently that *Dasein* is an entity that "must be existentially". In a second step, Weberman analyses the existential character of certain emotions, showing also other moods than *Angst* to be existentially relevant. As a conclusion, the paper suggests an adjectival use of the term existential such that it can also describe other things, e.g. artworks or experiences.

Lore Hühn examines in her paper *Com-passion* how Schopenhauer, relying on the essential identity of all living creatures, casts doubts on the primacy of reason in delivering a foundation of morals. Instead of reason, Schopenhauer argues, the capacity to suffer should be taken as the basis of an alternative model of ethics of compassion to an ethics of recognition. Furthermore, compassion is a distinguished experience that strikes the subject in his innermost core, for it concerns the subject's fragility and vulnerability as basic elements of its finitude. Hühn explores the theoretical proposal of an ethics of compassion critically, and concludes that the fundamental contradiction of an ethics of this kind is exhibited in the figure of the ascetic. She highlights that the sense of release (*Gelassenheit*) implied by the negation of the will excludes the normative reference to the other which for Schopenhauer was earlier the chief motive of moral action.

Hye Young Kim analyses the general characteristics of emotions like angst, guilt, fear, concern, and shame that are regularly treated in Existentialism. In her view, these emotions are existentially relevant, because they belong to the core of human existence in its finitude. In addition, Kim underlines the Christian theological element in the interpretation of human existence. The paper investigates other emotions rather neglected in these discussions such as gratitude and forgiveness, and makes a case for the claim that these emotions are fundamentally related to the understanding of human existence.

James Cartlidge's paper articulates doubts whether Martin Heidegger's frequent refusal of the categorization of *Being and Time* as philosophical anthropology is justified. Cartlidge finds Heidegger's argument that his project as 'fundamental ontology' cannot be a piece of philosophical anthropology is not convincing, since at the very heart of Heidegger's project is an analysis of the structures of the existence of '*Dasein*'. Despite Heidegger's all protestations, *Dasein* is an entity that human beings are an instantiation of, the entity that has

a relationship of concern towards its existence and which is capable of raising the question of the meaning of Being. Cartlidge provides a sketch of philosophical anthropology as an attempt to understand what is common to all instances of human existence with its significant features and structures. He examines Heidegger's analysis of moods to show that his work is best understood as involving a kind of philosophical anthropology.

Philippe Cabestan discusses Freud's legacy, especially the hypothesis of the unconscious, with regard to its credibility. To do this, he first follows Heidegger's criticism based on the distinction of natural phenomena and human phenomena. In a second step, Cabestan considers Jean-Paul Sartre's concept of bad faith, because Sartre thinks that, for instance, the hysteric is aware of what he doesn't want to be aware of and, as long as he tries to escape from it, he is necessarily aware of what he tries to escape from. The paper argues that the concept of bad faith alone is not able to explain unconscious behaviour, and the theory of the unconscious needs to be liberated from tendencies that treat it as a noun (MacIntyre) or as a thing-in-itself (Sartre).

Csaba Olay examines three paradigmatic thinkers of alienation – Rousseau, Marx, and Lukács – in order to show a general structural problem of different conceptions of alienation. He identifies in Rousseau what might be called a simplified precursor conception of alienation which has the structure of possession and subsequent disappropriation of man's original constitution. The paper compares this view with a more specific version of alienation in Marx' thought that might be described with the possession – disappropriation – reappropriation formula. Olay analyses Lukács's critique of capitalist society within the Marxist tradition with an eye on how the concept of reification partly carries on and partly modifies the conception of alienated labour as a basic tenet of Marx's thought. The paper shows that Lukács could not clarify how non-alienated conditions should be conceived.

Philippe Höfele's paper seeks an evaluation of emerging technologies on the basis of Hans Jonas' "heuristics of fear" that constitutes a principle and a method for assessing new technologies without knowledge of their future consequences. Höfele shows that this "heuristics" offers more than assessing the risk of technical developments, since the fear reveals at the same time *ex negativo* what constitutes human existence as such. In Jonas' view, a new technology always appears in the self-image of mankind what is illustrated by his historical reference point in Heidegger's analysis of *Angst*. Heidegger's description clarifies the importance of preserving the horizon of possibilities for human *Dasein*. Höfele argues that Jonas completes Heidegger's analysis of the open character of existence with the proposal of a "selfless fear" that involves a collective We and future generations as well.

*Gergely Ambrus – Csaba Olay*



## Narrative Identity in its Crises in Modern Literature\*

### I. IN THE BEGINNING: TWO DECOMPOSITIONAL APPROACHES TO THE HUMAN SELF NOTION

The modern romance of narrative theories and selfhood goes back to an initial dual attitude in European modernity of anchoring our self-related notions into human experience. Both created alternatives to the stability and indivisibility of the Cartesian Ego as a starting point. One started from body image centered conceptions, being *centripetal* in this sense, while the other one started from the role relations of persons, and was *centrifugal* in this regard as summarized in Table 1.

*Table 1 Two decompositional theories of the self*

<i>Centripetal</i>	<i>Centrifugal</i>
Inner Ego: essence is inner coherence	Outer ego: consensual validation
Starting to build from body image	Starting from interaction
Constructing objects and reference frames	Constructing role repertoires
Condillac, Mach, Head	G. H. Mead, Vygotsky Tomasello, Gergely – Csibra

In the 20<sup>th</sup> century, due to moves in empirical psychology and philosophical and literary theories, an especially victorious version of centrifugal theories would identify the ‘outer layers’ of selfhood as patterns of stories tolled to others and to ourselves. This narrative turn was embedded within psychology first into issues of memory schematization.

\* This paper is based on two talks of mine. The first one with the above title was presented at the conference on *Narrative Identity and Narrative Understanding*, Eötvös University, Budapest, May 3<sup>rd</sup>, 2019, organized by Gergely Ambrus. The second one was given at the *Wiener Sprachgesellschaft*, in Vienna, January 21<sup>st</sup>, 2020, with the title *From experimental studies of story organization to narrative theories of Self*, on the invitation of Wolfgang U. Dressler. Suggestions from Paula Fisher, Bálint Forgács, Hanna Marnó and Kristóf Nyíri are highly appreciated, but not always accepted.

## II. RELATING STORIES TO THE NOTION OF SELF

Narratives, as we call them today, have become central to psychology as part of the general efforts towards a more meaning and schema, rather than association-based theory of memory in the mid-20<sup>th</sup> century. The main actor, Frederick Bartlett (1932), had shown that our understanding and memory processes are always contextualized. Recall is not a passive process, but a result of active schematization as his monographer Wagoner (2017) analyzed in detail. From the 1970s on, the schema theory of Bartlett was rediscovered as part of the ancestry of modern cyclic schema theories (Rumelhart 1980). One trend of these new schema theories used story like narrative materials. The narrative pattern ideas were imported to contemporary cognitive psychology from other social sciences, from folklore, from anthropology and literary studies, and while they infiltrated psychology, they soon reached a level of generality touching upon philosophical issues such as the use of anthropomorphic schemata, and the relationships between story telling practices and our naive notions of Self (Pléh 2020). The re-discovery of the Russian folklorist Vladimir Propp (1928/1958) was a central step in this process. Propp was working at the same time Bartlett was experimenting with his diffusionist ideas in story schematization. Propp realized that strict rules or regularities are hidden behind the fantasy-rich world of our European folktale heritage. Folktales have a skeletal underlying structure, and they are characterized by a limited number of ‘roles’ and ‘functions.’ We can see these ‘roles’ in recent cultural theories as special attractors that are related to our folk psychology notions regarding human agency and its underlying motivations. This attitude was rediscovered and taken over into modern narrative research by Colby (1973) analyzing the corpus of Eskimo folktales. Colby modernized the conceptual approach and proposed a generative grammar for the corpus of Eskimo tales. From repeated patterns, we extract schemata and templates – among them the story templates – and use these to interpret new events. The construction of cognitive templates is based both on subliminal perceptions of human life and on experience with the array of cultural models available. The cultural models themselves, being patterned and ‘ready-made’ in a coded, condensed form, yield information for the anthropologist on the nature of these templates (Colby 1966). Thus, in this vision, story structure tells us about the structure of mundane social reality and the place of self in it (Colby 1966, 1975).

Dozens of cognitive psychologists starting from Rumelhart (1975) have taken up these ideas to see how can we build story grammars and how they help to operationalize the concept of schemata. After a short excursion into strictly formal models, these efforts turned to theories of naive social psychology, specifically theories of attributed intentional action as the explanation for schematization effects. In direct comparison, the predictions derived from the Schank and Abelson (1977) Causal Chain model had the most explanatory power in predicting

recall patterns (Black and Bower 1980, Pléh 1987). Graesser (1996) showed that in text understanding the causal and intentional naive attribution models are used in a complementary manner, and for human actions, we use an intentional frame. In understanding and recalling stories, we mobilize our naive social psychology about the structure of human action and about the usual motives for action. The coherence is found by the hearer–reader through the projection of these motivated action schemata to the story (Pléh 1987, 2003, László 2014).

Stories as a special type of narration require a hero, who has a system of goals, as well as a perspective. The hidden coherence of stories is provided by the problem-solving path of the hero (Black and Bower 1980) within a motivational field that is created by the goal system of the hero, such as the motives of hunger and the like, as Bartlett (1923, 1925) was already aware of.

There was an interesting meeting of paradigms when the structure-hungry cognitive psychologists themselves had to turn to theories (and even naïve, folk) theories of human action to account for what Bartlett labeled as schematization. A search for coherence underlies our schematization of stories, and this coherence is basically found by “turning on” our machinery of intentional attributions, and thereby reconstructing a causal chain that consists of causes and reasons that lead to these events.

### III. SELFHOOD IN ELABORATED NARRATIVE THEORIES

The entire notion of schematization and the uses of stories to prove it (narrativity) have already suggested for Frederic Bartlett (1935. 311) a constructive approach to the issue of self as well. “There may be a substantial Self, but this cannot be established by experiments on individual and social recall, or by any amount of reflection on the results of such experiments.”

In contemporary psychology there was a move towards interpreting narrative schematization as based on the use of the naive theories of intentional action. Parallel to this development, there were moves in three domains towards a more elaborate narrative interpretation of the Self.

- The theory of narrative and descriptive knowledge forms proposed by Bruner.
- Philosophical theories of narrative self by Dennett and Ricoeur.
- Literary theories on the relations between the modern novel and the modern Self.

### 1. *Narratives as primary organizations of knowledge (Bruner)*

The narrative approaches in contemporary psychology show up as *flexible models of the world opposed to essentialism*, as phrased by one of the leaders of the new movement, Jerome Bruner (1987, 1990, 1991). Essentialisms in this sense relates to the idea of a stable Cartesian Ego. The new model of the world contrasted with this in psychological narrative theories consists of a socially constructed world, and a socially constructed Ego, where the work of our self-narratives, or life stories, would be central to this constructive process. Bruner postulated two basic different approaches to the world. There is an intention and goal-based narrative, and a descriptive agentless approach to the world.

The duality shown in Table 2 gives an interpretation regarding the classical hermeneutic and causal duality dividing psychology and gives a primacy for narratives. Narratives treat events in an anthropomorphic way, in this regard hermeneutically.

*Table 2 The narrative and paradigmatic modes of cognition proposed by Bruner (1985, 1990)*

<i>Cognitive mode</i>	<i>Narrative</i>	<i>Paradigmatic/descriptive</i>
Organization	temporal, sequential, human action based	timeless, categorical, logical (Platonic) hierarchical
Discourse types	story: intentional teleology	description: relationships
Ideals	uniqueness, episodes	impersonal validity
Embedding	context: personal and social	decontextualization
Disciplines	humanities	sciences

In the vision of Bruner, children are attuned to these two ways of organizing knowledge. The narrative one is the personal, the paradigmatic (descriptive, theoretical) one is the categorical, scientific one. The narrative approach is always more primary, more elementaristic, and more readily available. This is the primary way to approach anything to make it meaningful. The Hungarian social psychologist János László (1999) pointed out that these attitudes do vary even within a single culture: we can approach, for example, a historical event as an embodiment of categories, and as a representation of individual fates and events.

Narrative metatheory assumes that the coherence of our internal world is provided by storytelling. There is a further question regarding the origins of these interpretation patterns. The initial questions regarding what gives pattern to simple stories, found an answer in “naive social psychology.” One has to somehow answer the following question: where do patterns of naive social psychology originate from? There are two rival solutions here. The modular vision would basically state that some kind of intentional and even teleological attribution is a modular feature of the human mind developing very early on (Csibra – Gergely 2007), and this makes our narrative patterns coherent. Bruner and his followers

would claim, on the other hand, that the naive teleological theory itself unfolds as a very result of experience with narrations (Bruner 1985, 1990, 1996), carrying a strong social emphasis about the origin of our attributing schemata.

From a developmental perspective, Bruner suggests that by distinguishing between outside (“real life”) events, the inner life of the hero, and the reactions of the narrator, storytelling practices foster the distinction between objective reality and mental reality. This aspect of stories has the challenging implication that narration is somehow intimately tied to our models of personhood and self as well. The world of narration would be making the connection between the real world and our inner world (our Self). Narratives provide us with perspectives to help to “give meaning” to whatever happens to us (Bruner – Lucariello 1989).

The concentration on actual stories as intellectual and cognitive organizing tools as interpreted by Bruner (1985, 1987, 1997), has become part of the modern anti-essentialist movement. Self as a safe Cartesian starting point and the world of stable objects is replaced by a world socially constructed through narratives and a Self that is as well constructed by narration. The world of narration relates the social world and our inner world. This bridging would be a crucial anchoring point for the centripetal, interactionist world view.

The difficulty lies in the fact that life is lived forward, encounter by encounter, but Self is constructed in retrospect, meta-cognitively. [...] Our self-concepts are enormously resilient, but as we have learned tragically in our times, they are also vulnerable. Perhaps it is this combination of properties that makes self-such an appropriate if unstable instrument in forming, maintaining, and assuring the adaptability of human culture. (Bruner 1997. 159.)

The theory of Bruner is rather abstract in itself and takes narratives as possible organizing tools of experience with no effort to operationalize these proposals. Several lines of research in developmental identity theory and clinical psychology tried to combine this theoretical narrative attitude with actual study of Self-related narratives. In this way, narratives as the basis of our notion of self started to be integrated into data on life stories and on autobiographical memory (McAdams 2001, McAdams and McLean 2013).

## *2. Philosophical narrative self theories*

The narrative trend also emerged as a philosophical proposal that makes *narratives essential for the organization for our notions of Selfhood*. These claims showed up in otherwise rather divergent, partly phenomenological, partly analytic theories (Ricoeur 1965, Taylor 1989, Dennett 1988, 1990, 1992).

From a phenomenological attitude, Ricoeur (1965) started off from a philosophical reinterpretation of psychoanalysis. The talk of the patient was no longer seen as a symptom of the unfolding of some internal essences, such as the natural processes of libidinal development, but as text, and he interpreted the work of the psychoanalyst similar to the work of a literary scholar, as text interpretation. In later elaborations of his narrative theory towards issues of identity, Ricoeur (2004) holds narrative identity responsible for mediating the two poles of personal identity, the pole of sameness (*idem*), referred to by what we call character, a set of innate or acquired attitudes and capacities, and the pole of selfhood (*ipse*), including trustworthiness and faithfulness to oneself, despite all the deviation and transformations which mark the path of the Self.

In the analytic corner of philosophy, Dennett (1991) in his anti-Cartesian view on consciousness and Selfhood – treating them in tandem – started from a narrative metatheory. Dennett basically claims for a soft and constructed theory of the Self.

A self, according to my theory, is not any old mathematical point, but an abstraction defined by the myriads of attributions and interpretations (including self-attributions and self-interpretations) that have composed the biography of the living body whose Center of Narrative Gravity it is. (Dennett 1991. 426–427.)

In the view of Dennett, there is no internal agent in a Cartesian Theater who would make things coherent. Coherence comes as a relaxation point in forging intentional sequences out of the events coming to us. We make Multiple Drafts of every incoming event (another narrative metaphor), and there is one of these that under normal circumstances is treated as being a conscious stage in information processing for the same sequence of events; that is, several “stories” are created.

The novelty of Dennett’s theory is twofold. For him, the level providing us with meaning and coherence, does not require a disembodied mind. This level is set into a narrative and intentional model that in principle will have an evolutionary story to it (Dennett 1994). Our self-notions are related to the fact that we are at the same time authors and audiences of our self-narratives. “People constantly tell themselves stories to make sense of their world, and they feature in the stories as a character, and that convenient but fictional character is the self” (Dennett 1992. 24).

### *3. Narrative selfhood and literary theory*

Both Ricoeur and Dennett made excursions in their narrative theories of the self toward literary narratives. In the mid-20<sup>th</sup> century, explicit theories of literature also spelled this out. The life philosophy embedded in classical narration is the

idea that there is a continuous, intelligible life with initiatives that is full of Plans. These Plans give coherence of the narrator and of narration. As Kundera, the Czech-French writer (1986. 58) expressed it: “Out of the mysterious and chaotic fabric of life, the old novelists tried to tease the thread of a limpid rationality; in their view, the rationally accessible motive gives birth to an act, and that act provokes another. An adventure is a luminously causal chain of acts.”

Seen from this perspective, traditional *narrative schemata* with their mobilization of intentional action interpreting modules are powerful coherence building devices. The specificity of traditional simple stories lies in the fact that due to the prototypical motivations in a given culture, and due to the simple transparent narrative point of view, this action organization can be revealed easily and unequivocally on the part of the understander. One of the clearest aspects of the transformation of these patterns in modern “high literature” concerns the changes in the comprehensive Plans of action from the point of view of the Hero and/or the Narrator. Its presence gives coherence to classical narratives, be it fairy tales – the youngest boy wants to marry a king’s daughter, sets out into the world, and through many obstacles gets her – or the bourgeois novel where the young hero comes to the big city, wants to make a career, relying on relatives and women reaches these goals. The comprehensive message of the work is tied to the intentional system of the hero (Pléh 2003, 2019).

Traditional European fiction has become a central effort towards this cultivation of Self through cultivation of narratives. As the writer and literary theorist David Lodge (1992, 2002) claimed in detail, the modern self and the modern novel were born together.

In the reading of novels, the already existing narrative self concept was individualized. The idea of the omnipotent writer developed together with the idea that there are three layers to a novel – the layers of external actions, internal plans, and feelings. The mutual relating of these three layers has provided for classical developmental novels and the integrity of the novel. Everything was seen in the unfolding of the hero. The unfolding of the hero gives a model for our own unfolding. (Pléh 2019. 245.)

This has interesting implications about the relativity of the narrative approach regarding the Self.

We have to acknowledge that the Western humanist concept of an independent individual self is not universal, not eternally valid for all places and times, but is a historical and cultural product. That does not necessarily mean that it was not a good idea and its time is over [...] We also have to acknowledge that the individual self is not a fixed and stable entity, but is always created and modified in our consciousness during interaction with others. (Lodge 2002. 91f.)

The entire issue of narrativity and the connections between self-narratives and the notion of Self has become central in the general cultural discussion regarding the “disappearing self”.

#### IV. DISSOLUTIONS OF THE SELF, THE MIND AND OF THE NOVEL

The issue of modern novel organization is the point where the narrative frame issue becomes intimately tied to the crisis of modernity and to the problem of the relations between the changes of narrative patterns and a crisis in our view of ourselves. There is a remarkable similarity in the way narratives become central in experimental psychology, in the study of development and in the cultural and philosophical theorizing about the centrality of narration in our self-image, as we have seen above. A similar affinity appears in issues of dissolution. Our present intellectual world in the early 21<sup>st</sup> century can be characterized by two types of dissolutions (or, if you prefer, crises). Similar crises went on several times during the 20<sup>th</sup> century. The first crisis is the dissolution of the stable Ego, which was already characteristic of the late 19<sup>th</sup>-century philosophy and psychology that became, with the words of the Hungarian philosopher Kristóf Nyíri (1992), “impressionistic” in its search for stable reference points.

The other, parallel dissolution or disintegration, went on in the realm of culture. One dominant aspect of this in the early 20<sup>th</sup> century was a dissolution of traditional patterns of narration. There are interesting parallels here between artistic practice and philosophy. Kristóf Nyíri (1992) analyzed the affinities between the elementaristic theory of mind proposed by Ernst Mach (1897), and the school of impressionistic painting. The strong drive to liberate yourself from anything secondary, knowledge-based (top-down), anything schematic, and a search for undeniable, original certainty lead to pictorial and epistemological impressionism: the real raw stuff of both would consist of patch-like pieces of experience. There was a similar trend in questioning the validity of traditional narrative schemata and the underlying naive application of the intentional stance to narrative agents as well. There are interesting parallels between giving up the idea of a causal chain in the outside social world of the novel and questioning the presence of an integrative Ego in the inner world of the novel (Kundera 1986).

As the Italian editor and cultural philosopher Scalfari (2012) pointed out, there were tensions in European criticisms after the great works of Tolstoy and Dostoevsky. Is the novel dead? These death calls were however followed by works of Marcel Proust, Joyce, and Kafka. “What was finished was the romantic and naturalist novel. The novel which described the bourgeoisie with its tropes, passions, hypocrisies and vices” (Scalfari 2012. 207).

In the new types of narrations taking shape in the 20<sup>th</sup> century, the godlike image of an author with all-encompassing knowledge is replaced by either a



direct presentation of the inner world, or with a description of external behavior with no pre-assigned perspectives. The great discovery of Proust was to turn towards the inner life. “The striking innovation was to accomplish a travel inside the self rather than in the social world of the times” continues (Scalfari 2012. 209), at the same time realizing that the essential point is the loss of the plot.

Narration dominated by the intentional stance in the sense of Dennett (1990) is replaced by a presentation of internal mosaics, which could already be observed in Virginia Woolf, Proust or Joyce, or half a century later, in the French *Nouveau Roman* and the French absurd, like Beckett or Ionesco. Likewise, this model of internal mosaics was also present in the theoreticians and practitioners of postmodern literatures. Woolf herself made the new ideas very provocative, referring many times to the writing practice of James Joyce and Proust (Lewis 2008). Writers spend too much time in recreating a plot. Virginia Woolf campaigned for a new style of writing. For her, “[to] provide a plot, to provide comedy, tragedy, love interest” is all artificial, and a tyrannical obedience to tradition (Woolf 1925. 160). If the

writer were not a slave but free [...] there would be no plot, no comedy, no tragedy, no love interest or catastrophe [...] let us record the atoms as they fall upon the mind in the order in which they fall. [...] the point of interest lies very likely in the dark places of psychology (Woolf 1925. 161, 162).

With the advent of the ‘no story stories,’ different versions of new narration emerged as variations on defocusing:

- we do not know who we are (Musil)
- we do not go anywhere (Camus)
- heroes are not lords of their fate (Kafka)
- heroes are slaves to forces beyond reach of their consciousness (Proust).

These changes of motivational structure went together with psychological defocusing from the clear differentiation of Internal Plans and External Actions.

- Dissolution into memory (Proust)
- Dissolution into stream of consciousness (Joyce)
- Dissolution of roles (Musil)
- Challenges to intentional action (Gide)
- Presenting only the behavioral skeleton (Hemingway)

With the birth of the modern novel in Proust, Joyce, Woolf, and Musil, writers show that Kundera is right in a central respect: modern writers were experimenting with knowledge structures, and they prefigured a narrative concept of identity (including all of its crises) well before it was formulated as a theory of mind by philosophers, and they give a rational reconstruction to the stream of

consciousness through narration. "All novels, of every age, are concerned with the enigma of the self" (Kundera 1986. 23).

Musil himself, as shown by Freed (2007), tried to combine his philosophical training and dissatisfaction with philosophy by a new writing technique. The essayistic writing was a way to find a compromise between the philosophical decomposition arriving with modernity, and the need for coherence. Samuel Beckett (1931), the later Nobel prize winner master of absurd writing, gives a similar interpretation of the importance of the multiple and non-conscious construction of the Self in Proust: "But here, in that 'gouffre interdit à nos sondes' is stored the essence of ourselves, the best of our many selves and their concretions that simplists call the world." (Beckett 1931. 31) Kundera (1986) sensitively presented how this type of goal-coherence was ruined in the novels and realities of Franz Kafka. The hero is subjected to non-transparent Plans of others, and these Plans do not become clear even till the end of the story. The continuous goal system disappeared before Kafka as well. It was replaced by the world of inner experience in Joyce, and in Proust, as analyzed by Beckett, the action-based logic of narration and the actions of the hero are replaced by an undifferentiated network of experience, imagery, and souvenirs.

The identification of immediate with past experience, the recurrence of past action or reaction in the present, amounts to a participation between the ideal and the real, imagination and direct apprehension, symbol and substance. Such participation frees the essential reality that is denied to the contemplative as to the active life. What is common to present and past is more essential than either taken separately. (Beckett 1931. 74.)

The comprehensive Plan disappears not only in the impressionistic presentation of the internal world but also on the level of behavior. In this third type of modern writing the external behavior is not characterized by clear Plans. Rather, things just happen to the hero, and he acts reactively, and tries to give meaning to the actions only afterwards.

The continuous world of intentions is replaced not by an inner world of experience, but by the world of behavior. Think of some of the acts of Mersault in *The Stranger*, of another Nobel prize winner, Albert Camus or to the beginning acts of the actor Belmondo in Godard's movie *À bout de souffle*. The reader and the viewer are immediately presented by pieces of behavior, without enough preparation for the setting, and without a possible intentional interpretation. The individual experiences and acts are not presented as parts of an encompassing Plan. They can only be given a local interpretation. He shot the cop asking for his papers, but this happened so fast that neither he (the hero, Belmondo), nor we, the viewers had any chance to build up a plan to motivate the deed (*À bout de souffle*). In a secondary way, we give interpretation to something that al-

ready happened. We make a story out of it like psychoanalysts, but the unique un-interpreted act preceded the story, while in classical narrative patterns, the starting point is the story with its intentional layout, and unique events fill the slots in a secondary way. Classical narration treated the narrative pattern in an essentialist way, with a belief in the integrative Self, and the events being only manifestations of this. This is in accordance with the top-down style of writing, and with the idea of an omnipotent and omniscient writer.

The key scene from *The Stranger* illustrates this lack of narrative build-up relying on intentions:

Then everything began to reel before my eyes, a fiery gust came from the sea, while the sky cracked in two, from end to end, and a great sheet of flame poured down through the rift. Every nerve in my body was a still spring, and my grip closed on the revolver. The trigger gave, and the smooth underbelly of the butt jugged my palm. And so, with that crisp, whipcrack sound, it all began. (Camus 1954. 76.)

The murder by Meursault is rather different from that of Lafcadio in the *Caves of the Vatican* by André Gide. His act (throwing a passenger out from the train cabin) is quoted as the classical example of *action gratuite*. This is an act of “no motive,” however, only in the sense of bringing no utility to the actor. Otherwise, Gide, writing in classical style, makes sure that we see it as a planned, intentional, premeditated action. Lafcadio even laughs in advance how much trouble the police will have in dealing with a *crime sans motif*, with an unmotivated crime.

It’s not so much about events that I’m curious, as about myself. There’s many a man thinks he is capable of anything, who draws back when it comes to the point... What a gulf between the imagination and the deed! [...] And no more right to take back one’s move than at chess. Pooh! if one could foresee all the risk, there’d be no interest in the game!” (Gide 1914/1953. 186.)

There are plenty of more recent examples for the dissolution of intentional coherence. In this respect, Christine Brooke-Rose (1986) presented an interesting outline for the changes in writing so typical of modern (and of course postmodern) literature. First came the defocalization of the hero. That was already present in the nineteenth century. Think of the well-known comparisons regarding the Waterloo battlefield descriptions by Victor Hugo in *Les Misérables*, where you have an epic enumeration combined with a panoramic view and a clear presentation of the scenery, with the scene of Fabricio del Dongo being part of the great battle in Stendhal’s *La Chartreuse de Parme* without really knowing it. The entire scene is defocused: we see the hero as being entirely out of the intentional plans of the agents, unaware of their plans, and even of them being agents. He

does not even realize he is seeing the great man he came for. He is part of the battle without knowing he is in Waterloo.

This is the defocusing of the intentional plans, indeed. This was further combined with a defocusing of the “survival value” of the hero. Present-day heroes are no more close friends of ours, as were Madame Bovary or Anna Karenina, or Rastignac, to that effect.

There are tragic, ambiguous, and ironical overtones as well in the literature regarding the dissolution of the Self. Both leave one central issue open, however. When we dissect the Self into elementary experiences and their relationships, and narration into narrative morsels, do we make them disappear by this very act? Does the Self really disappear, or do we only claim that compared to the primary stuff of experiences, it is only secondary? (That is the way, for example, how Beckett interpreted Proust.) Does narration disappear, or is it only a secondary organization compared to the primary thread of discourse? Do we manage to radically eliminate coherence, which is usually accounted for by the Self and by narration, or do we only make it secondary rather than using it as a starting point?

Narrative metatheory as a non-essentialist view of coherence takes the second option. Rather than postulating a substantial Self, the coherence of our internal world comes around by milder means, by storytelling.

The issue of coherence in communicative terms implies that the partners, A and B must follow a mutual, joint model. They must allow each other to reconstruct similar relationships between the individual propositions. This is referred to as the maxim of relevance by the communication model of Paul Grice (1975), as well as in the elaborated model of Sperber and Wilson (1995) and as the issue of higher-order models of intentionality by Dennett (1987).

The concept of communicative coherence allows us to look for inner coherence in a non-essentialist way and to avoid the usual pendulum-like shifts between disintegrated and essentialist views of the Self. The notion of coherence might be a help in finding some peace in the chaos of the world, without necessarily committing ourselves to another round of essentialism. As the new theory promoted by Mercier and Sperber (2017) claims, even human reasoning should be interpreted as an evolved tool of making arguments. Narrative patterns in this sense would be coherence building devices, with a high attraction value (Boyd 2009), a peculiar type of argumentation based on cultural expectations.

The system proposed by Daniel Dennett (1987, 1991, 1992, 1994) has some intellectual promise here, especially since he consciously connects the two types of dissolutions, that of the Self and of the narrative patterns. For postulating coherence, he does not need a hypostasized subject. The coherence of inner life (the Self, if you like) will be a “soft notion” for him, a “gravitational point” of all the stories we tell ourselves. It is interesting to see that the dynamic nature of consciousness, and the multiple nature of self, introduced more than a

century ago by James (1890) as a response to the crisis of *fin de siècle* society, and as an application of the evolutionist non-essentialist image of inner life, comes back in different forms over the entire century. The narrative turn is connecting the original association with the stream of consciousness idea with new ways of writing.

These efforts may not bring happiness over the issue of the disappearing Self, but they may contribute to more sensible interactions between philosophy and cognitive sciences. As Gallagher put it in a programmatic survey:

By extending the ideas of a narrative self, we are perhaps coming closer to a concept of the self that can account for the findings of the cognitive sciences and neurosciences, as well as our own experience of what it is to be a continuous, phenomenological self. [...] Philosophical ideas about the self can be aligned with, and can inform, current ideas in cognitive science. I also believe that philosophers can learn about the nature of the self from psychologists, neuro-scientists and other cognitive scientists. Thus, collaborative efforts between philosophers and scientists promise to open up more subtle and sophisticated avenues of research, which will define more fully the concept of the self. (Gallagher 2000. 20.)

## REFERENCES

- Bartlett, Frederic C. 1920. Psychology in Relation to the Popular Story. *Folklore*. 31. 264–293.
- Bartlett, Frederic C. 1923. *Psychology and Primitive Culture*. Cambridge, Cambridge University Press.
- Bartlett, Frederic C. 1932. *Remembering*. Cambridge, Cambridge University Press.
- Bartlett, Frederic C. 1932. *Remembering. A Study in Experimental and Social Psychology*. Cambridge, Cambridge University Press.
- Beckett, Samuel 1931. *Proust*. Cited after the edition of 1987. London, J. Calder.
- Black, John B. – Gordon H. Bower 1980. Story Understanding as Problem Solving. *Poetics*. 9. 223–250.
- Boyd, Brian 2009. *On the Origin of Stories. Evolution, Cognition, Fiction*. Cambridge/MA, Harvard University Press.
- Brooke-Rose, Christine 1986. The Dissolution of Character in the Novel. In Thomas C. Heller – Morton Sosna – David E. Wellbery, (eds.) *Reconstructing Individualism: Autonomy, Individuality, and the Self in Western Thought*. Stanford/CA, Stanford University Press.
- Bruner, Jerome 1985. *Actual Minds and Possible Worlds*. Cambridge/MA, Harvard University Press.
- Bruner, Jerome 1987. Life as Narrative. *Social Research*. 54. 11–32.
- Bruner, Jerome 1990. *Acts of Meaning*. Cambridge/MA, Harvard University Press.
- Bruner, Jerome 1991. The Narrative Construction of Reality. *Critical Inquiry*. 18. 1–21.
- Bruner, Jerome 1996. *The Culture of Education*. Cambridge/MA, Harvard University Press.
- Bruner, Jerome 1997. A Narrative Model of Self-Construction. *Ans NY Acad Sciences*. 881. 145–161.
- Bruner, Jerome 2006. La culture, l'esprit, les récits. *Enfance*. 58. 118–125.
- Bruner, Jerome 2008. Culture and Mind: Their Fruitful Incommensurability. *Ethos*. 36. 29–45.

- Bruner, Jerome – Joan Lucariello 1989. Monologues as a Narrative Recreation of the World. In Katherine Nelson (ed.) *Narratives from the Crib*. Cambridge, Harvard University Press. 234–308.
- Camus, Albert 1954. *The Stranger*. Transl. by Stuart Gilbert. New York, Vintage Books.
- Colby, Benjamin N. 1966. Cultural Patterns in Narrative. *Science*. 151. 793–798.
- Colby, Benjamin N. 1973. A Partial Grammar of Eskimo Folktales. *American Anthropologist*. 75. 645–662.
- Colby, Benjamin N. 1975. Culture grammars. *Science*. 187. 913–991.
- Csibra, Gergely – György Gergely 2007. ‘Obsessed with Goals.’ Functions and Mechanism of Teleological Interpretation of Actions in Humans. *Acta Psychologica*. 124. 60–78.
- Dennett, Daniel 1987. *The Intentional Stance*. Cambridge, MIT Press.
- Dennett, Daniel 1989. The Origins of Selves. *Cogito*. 3. 163–173.
- Dennett, Daniel 1990. The Interpretation of Texts, People and Other Artifacts. *Philosophy and Phenomenological Research*. 1. Supplement. 177–194.
- Dennett, Daniel 1991. *Consciousness Explained*. Boston, Little Brown.
- Dennett, Daniel 1992. The Self as a Center of Narrative Gravity. In Frank S. Kessel – Pamela M. Cole – Dale L. Johnson (eds.) *Self and Consciousness: Multiple Perspectives*. Hillsdale/NJ, Erlbaum.
- Dennett, Daniel 1994. *Darwin’s Dangerous Idea*. New York, Simon and Schuster.
- Freed, Mark M. 2007. Robert Musil’s Other Postmodernism: Essayism, Textual Subjectivity, and the Philosophical Discourse of Modernity. *Comparative Literature Studies*. 44. 231–253.
- Gallagher, Shaun 2000. Philosophical Conceptions of the Self: Implications for Cognitive Science. *Trends in Cognitive Sciences*. 4. 14–21.
- Gaesser, Arthur C. 1996. *Models of Understanding Text*. Mahwah/NJ, Lawrence Erlbaum.
- Grice, Paul 1975. Logic and conversation. In Peter Cole – Jerry L. Morgan (eds.) *Syntax and Semantics*. Vol. 3. *Speech Acts*. New York, Academic Press.
- James, William 1890. *Principles of Psychology*. New York, Holt. Digitized version <http://www.bahaistudies.net/asma/principlesofpsychology.pdf> or <https://archive.org/details/theprinciplesofp01jameuoft/page/n6/mode/2up>.
- Kundera, Milan 1986. *The Art of the Novel*. New York, Grove Press.
- László, János 1999. *Cognition and Representation in Literature. The Psychology of Literary Narratives*. Budapest, Akadémiai.
- László, János 2014. *Historical Tales and National Identity: An Introduction to Narrative Social Psychology*. Routledge, London.
- Lewis, Pericles 2008. Proust, Woolf, and Modern Fiction. *The Romantic Review*. 99. 77–86.
- Lodge, David 1992. *The Art of Fiction*. London, Penguin Books.
- Lodge, David 2002. *Consciousness and the Novel*. London, Penguin Books.
- Lodge, David 2009. The Best Stream-of-Consciousness Novels. *The Guardian*. Tuesday, 20 January.
- Mach, Ernst 1897. *Contributions to the Analysis of Sensations*. Trans. C. M. Williams. Chicago Open Court.
- McAdams, Dan P. 2001. The Psychology of Life Stories. *Review of General Psychology*. 5. 100–122.
- McAdams, Dan P. – Kate C. McLean 2013. Narrative Identity. *Current Directions in Psychological Science*. 22. 233–238.
- Mercier, Hugo – Dan Sperber 2017. *The Enigma of Reason*. Harvard University Press.
- Nyíri, Christoph J. 1992. *Tradition and Individuality*. Dordrecht, Kluwer.

- Pléh, Csaba 1987. On Formal- and Content-Based Models of Story Memory. In L. Halász (ed.) *Literary Discourse. Aspects of Cognitive and Social Psychological Approaches*. Berlin, DeGruyter. 100–112.
- Pléh, Csaba 2003. Narrativity in Text Construction and Self Construction. *Neohelicon*. 30. 187–205.
- Pléh, Csaba 2019. Narrative Psychology as Cultural Psychology. In Gordana Jovanović – Lars Allolio-Näcke – Carl Ratner (eds.) *The Challenges of Cultural Psychology Historical Legacies and Future Responsibilities*. London, Routledge. 237–249.
- Pléh, Csaba 2020. From the Constructive Memory of Bartlett to Narrative Theories of Social. *Culture and Psychology*. 26. 287–299.
- Propp, Vladimir 1928/1958. *Morphology of the Folktale*. Bloomington, Indiana University Press.
- Ricoeur, Paul. 1965. *De l'interprétation. Essai sur Sigmund Freud*. Paris, Le Seuil. In English *Freud and Philosophy An Essay on Interpretation* (1970.) New Haven, Yale University Press.
- Ricoeur, Paul 1990. *Soi-même comme un autre*. Paris, Seuil.
- Ricoeur, Paul 2004. *Memory, History, Forgetting*. Chicago, University of Chicago Press.
- Rumelhart, David 1975. Notes on a Schema for Stories. In Daniel G. Bobrow – Allen Collins (eds.) *Representation and Understanding*. New York, Academic Press. 211–236.
- Rumelhart, David 1980. Schemata: The Building Blocks of Cognition. In Rand J. Spiro – Bertram C. Bruce – William F. Brewer (eds.) *Theoretical Issues in Reading Comprehension*. Hillsdale/NJ, Erlbaum. 33–58.
- Scaffari, Eugenio 2012. *Par la haute mer ouverte. Notes de lecture d'un Moderne*. Paris, Gallimard.
- Schank, Roger C. – Robert P. Abelson 1977. *Scripts, Plans, Goals, and Understanding: An Inquiry into Human knowledge Structures*. New York, Halsted.
- Sperber, Dan – Deirdre Wilson 1995. *Relevance: Communication and Cognition*. Second Edition. Oxford/Cambridge, Blackwell Publishers.
- Taylor, Charles 1989. *Sources of the Self: The Making of Modern Identity*. Cambridge/MA, Harvard University Press.
- Wagoner, Brady (ed.) 2017. *Handbook of Culture and Memory*. Oxford, Oxford University Press.
- Wagoner, Brady 2017. *The Constructive Mind. Bartlett's Psychology in Reconstruction*. Cambridge, Cambridge University Press.
- Woolf, Virginia 1925. Modern Fiction. In Andrew McNeille (ed.) *The Essays of Virginia Woolf. Volume 4. 1925 to 1928*. London, The Hogarth Press. 157–164.



GÁBOR BOROS

## First Phase of the Narrative Theory of Personal Identity: Wilhelm Dilthey, and Georg Misch\*

In the last decades, a number of philosophical and psychological theories have made serious attempts to discover and make use of various aspects of different types of narratives from Proust's and Thomas Mann's "novels of time" through biographies and autobiographies to interviews with members of contemporary groups or individuals in therapeutic analysis or other particular situations. Their aims were not so much to make explicit the hidden linguistic structures of narratives but rather to understand identity in a broad sense, personal, group-, national, emotional, and other types of it. The initiators and proponents of these theories rarely referred to the two German philosophers, active between the mid-19<sup>th</sup> and the mid-20<sup>th</sup> century who attributed a fundamental role to autobiography as a particular kind of narrative both in history and in philosophy. In spite of their being neglected in this way, Wilhelm Dilthey and Georg Misch merit to be integrated in the history of the narrative identity-movement. For they emphasised much more than other theoreticians the fundamental importance of establishing narrative connections between the seemingly isolated events in life histories as the most effective instrument to establish meaningful and coherent life-units. From this perspective, it is promising to regard Dilthey and Misch as our contemporaries and to weigh up their contributions to a renewal and enrichment of the theory of narrative identity and the narrative theories of emotions. Within the framework of this paper, however, my modest aim is to persuade the benevolent reader that it is worth involving them in the general discourse on identity and narrativity.

The expression "first phase" in the title of this paper does not only mean chronological but also systematic priority. This is, however, far from being a matter of course. Dilthey and Misch did not publish works including the key words of contemporary narrative theories. Still, Dilthey opened up a path to approach through narrativity the discourse on self-understanding, self-interpretation as a positive result of his obstinately made attempts to get to grips with the problem

\* Special thanks to the NKFIH research projects K 120375 and K 129261 for their support during the period working on this paper.



of objectivity in what can be called human or historical sciences. My proposal is to interpret some cardinal passages in Dilthey's texts as starting points in a line of thought issuing in the great emphasis on autobiography in his late fragments. His pupil and son-in-law Georg Misch developed it through his life-long efforts into a monumental series of historical volumes on autobiography. At the same time, this development did not only consummate Dilthey's original concept. It also implied a gradual shift from the systematic, philosophical-gnoseological concerns of Dilthey to practical-historical ones. In my view, this latter development can fruitfully be connected to the philosophical-methodological attempts of some chief representatives of late 20<sup>th</sup> century human sciences to understand the methodological bases of their proper disciplines. Thus after a detour through narrative theory, one can regain some access to methodological questions.

My starting point is the attempt to emphasise the systematic-philosophical relevance of some important passages from Dilthey's fragmentary *Drafts for a Critique of Historical Reason*. Let us consider the following passage first. "The lived experience (*das Erleben*) is a temporal sequence in which every state is in flux before it can become a distinct object." (Dilthey 2002. 216.)

This seemingly simple sentence is a concise description of what we can certainly interpret as the *systematically* first moment when out of the unstoppable flux of life a complex, rudimentary mental phenomenon – "the lived experience" – shines forth, i.e. makes itself perceived. This *Erleben* is rudimentary and obscure because it does not yet contain "distinct objects" – and obviously, no distinct subjects either. This phenomenon is the nucleus of what Hobbes famously called the most miraculous among the phenomena of nature: "*to phainesthai*" – shining forth – itself. (Cf. *De corpore*, chapter 25; Hobbes 1996.) The sentence implies that in the first layer of the original flux of life the awareness or consciousness is – logically at least – missing. This is a flux of bodies mutually influencing each other; a flux that also includes everything that takes place in the human brain. This self-sustaining causal chain is the object of the physical-physiological viz. the neuro-sciences without, however, their having the slightest chance to tackle the nucleus of mental-spiritual life as such.<sup>1</sup> The mental-spiritual life is an autonomous layer of its own superimposed onto the equally autonomous layer of the corporeal.

The life of spirit manifests itself on the base of what is physical and represents the highest evolutionary stage on earth. The conditions under which the life of spirit emerges are developed by natural science in that it discovers a lawful order in physical phenomena. (Dilthey 2002. 217.)

<sup>1</sup> This does not exclude that they can find the bodily correlates of the mental phenomena or that the scientists proclaim that they identified the mental phenomena with some bodies.

One can, in fact, go a step forward and maintain that natural scientists can solely talk about “physical *phenomena*” because they apply precisely *their own* irreducible mental-spiritual faculties to apprehend bodily phenomena in physical sciences. These sciences are based on much more refined “life experiences” than the above mentioned nucleus is. In these experiences or rather experiments brought into laboratories or other artificial environments there are already distinct subjects and objects. Nevertheless, howsoever sophisticated the laboratory settings in these sciences may be, the emergence and elementary accomplishments of the life of the mind or spirit cannot be explained by physical-physiological-neurological sciences but must be accounted for on a higher level. “With lived experience we move from the world of physical phenomena into the realm of spiritual reality [...] The cognitive value of this realm is fully independent of the study of their physical conditions.” (Dilthey 2002. 217.)

“Lived experience” is, therefore, the interface in which the flux of the physical and the life of spirit are connected – a sublime version of Descartes’ pineal gland and Pufendorf’s divine creation of *ens moralia* above *ens physica*.

A step further, we can start with Dilthey reconstructing the processes of the mind. “Lived experience encompasses elementary operations of thought” (Dilthey 2002. 218). In this context, “elementary” means the moment of the go-between, as it were. It is not a particle or movement of a pre-given corporeal *substance*. Instead, it is the act of *giving*, a self-performing act of spirit, the coming into being of the kind of life superimposed onto the base of the physical; a *suprastance*, instead of a *substance*, as it were. “These operations occur when consciousness is intensified” (Dilthey 2002. 218).

This assertion obviously presupposes that there is already an original consciousness to be intensified. I interpret this presupposed first nucleus of consciousness as the one Spinoza describes in a difficult set of “axioms” as an alternative to Descartes’ “I think” (*cogito*) considered independent, and really distinct from all bodily processes. “Man thinks” [...] “We feel that a certain body is affected in many ways.” (Axiom 2 & Axiom 4 of *Ethics* 2, Spinoza 1985. 448.)

The two axioms read together point out a double cognitive state. “Man thinks”: in this first layer, there is neither a distinct subject nor a distinct object. This elementary indiscriminate cognitive state is followed by a state in which one can detect the base of what is “to be intensified” into the discriminate consciousness of the I-subject and the external affecting beings, the “distinct objects”.

After this short digression on Spinoza’s axioms with hints at the development of a discriminate cognition, we can now interpret more easily not only Dilthey’s above sentence but also the following one. “A change in a mental state of affairs leads to the consciousness of difference. In that which changes an isolated state is apprehended. (*An dem, was sich ändert, wird ein Tatbestand isoliert aufgefaßt.*)” (Dilthey 2002. 218.)

If we apply to this sentence the above interpretation of Spinoza's axioms, we can consider the indiscriminate "Man thinks" an indifferent "mental state of affairs" that opens up the way to the "consciousness of difference" in which a certain *enduring, identical, affected* body, and the continuously changing multitude of the *affecting* bodies are differentiated.

Equipped with this framework gained by way of the parallel interpretation of Spinoza and Dilthey, we can return to the simple sentence in our first quotation from Dilthey. The complex phenomenon that shines forth is consciousness in general borne in the dynamic relationship between an individual person as the bearer and a series of individual beings as the intentional objects of consciousness. On the object-side, this is a passive state: *being elevated into consciousness*, whereas on the person-side, the same event is a hardly separable mixture of activity and passivity, elevating and being elevated at once.

At this junction, the path also opens up that leads to giving an account of the linguistic means that express the experienced cognitive relations in the form of judgements. "Experiencing is followed by judgements that objectify what has been experienced" (Dilthey 2002. 218).

The basic building blocks of the logical structure of reality based on judgements also originate in the same elementary accomplishments of mind-spirit: they are called "formal" and "real categories" respectively: "[...] the formal categories spring from the elementary operations of thought. [...] such concepts are unity, plurality, identity, difference, degree and relation. They are attributes of the whole of reality." (Dilthey 2002. 218.)

The formal categories of "unity" and "identity" are destined to play an important role concerning autobiographies. The present quotation ends with a fragmentary beginning of a sentence: "The real categories..." Fortunately, however, this does not imply, that we are left at a loss when trying to account for the real categories. An earlier passage sheds some light on this issue: as we have seen, formal categories apply to the whole of reality, physical and mental, whereas real categories are to be employed as instruments for analysing the life of the spirit.

Among the real categories are those that originate in the apprehension of the world of human spirit [...] the life of spirit can everywhere be characterized in terms of productive systems, force, value, etc.

Temporality is contained in life as its first categorical determination and the one that is fundamental for all the others. (Dilthey 2002. 214.)

If we imagine a being in a timeless eternal now it will hardly have anything to communicate in a narrative form. But in reverse analogy, it is obvious that *temporality* is not only fundamental for every other real category but also a basic category for the human being as a being in time the identity of which is construed in and through the various narratives designed by herself or others about herself or

others; this latter is necessary, not the least because the finite consciousness that have been awakened through the above steps cannot exist in solitude without communicating its contents. A necessary condition of its coming into existence is that the being that is to become conscious is perceiving itself as the identical subject opposed to the unstoppable flux of objects affecting it in many ways. It perceives itself as the faculty of connecting a plurality of states already in the sense of its being the unique subject of a series of joint affections, mostly passive, rarely active. But more importantly, it is a more and more *active* pole generating connections in at least two senses. The first sense is the basic level experience of plurality without making any judgements, the second is the transformation of the basic indiscriminate feelings in judgments in which the perceiving subject “objectifies” (*vergegenständlicht*) the *felt* reality into objects and properties that can be predicated of them. This insight can be made the point of departure for a special hermeneutical logic of life as it is exemplified in works of Georg Misch and Joseph König (cf. Misch 1994; König 1937).

But from the point of view of this paper, it is not the most important feature of Dilthey’s account of the emergence of consciousness amidst the stream of life. From this point of view, it is more important to remark, on the one hand that the distinguished role temporality receives in Dilthey could very well be used as the point of departure for both Heidegger in *Being and Time* to connect Husserl’s analyses with Dilthey’s and von Wartenburg’s, and Misch’s assessment to Heidegger’s attempt in Misch 1931/1967. On the other hand, it is important to connect Dilthey’s formal category of identity with the overarching real category of temporality in order to discover the particular meaning of the traditional hermeneutical relationship between the parts and the whole that it solely acquires in the sciences that have human life as their proper object, the sciences of mind or spirit.

[The formal category of the relation between whole and part] first acquires its own meaning in the realm of the human sciences from the nature of life and the understanding appropriate to it, namely, that of a nexus in which the parts are interconnected (Dilthey 2002. 219).

In Dilthey’s view summarised in this short quotation, human life has as its essential ingredient self-understanding, self-interpretation that fulfils its task through the connection of the parts in time to create wholes in form of coherent, meaningful temporal sequences. Meaning itself originates in such particular wholes.

Thus essential to life is that it grasps itself by way of connecting its moments as cohering parts of a whole. Yet, this whole cannot become identical with the by definition unperceivable original whole in and as the flux of life. Therefore, the connections within the whole are not naturally given but construed on the basis of a narratable coherence of the moments in time that are constituted *as* parts only when they are connected and so related to the meaning establishing whole.

*The hermeneutical circle in form of the dialectic of parts and wholes appears as the basic structure of the finite human life of mind-spirit.*

Departing from this first result, the analysis can be continued in two directions.

The one is to draw the outlines of the theory of emotions to be based on the hermeneutical theory of the “elementary operations” of consciousness and the dialectic of parts and wholes in narratives. It seems that at least the most important human emotions are to be construed in a hermeneutic-holistic manner as consisting of parts that cannot be conceived *as parts* before relating them to the whole and *vice versa*. Relying on certain insights from Wittgenstein, P. Goldie applied this view to show in which way his version of a narrative theory of emotions could be built up (cf. Goldie 2014). I myself have tried to proceed further in this direction in recent texts of mine in which I termed the meaning establishing coherence of the parts a *narrabile* (cf. Boros 2017). The *narrabile* is in continual change, and so can and must be narrated again and again at least for ourselves to support our claim to be an identical person by way of a reassuringly meaningful narrative. In a forthcoming article I am planning to unfold this germ even further.

The other direction is to unfold the implications of the circular character of the constitution of a meaningful life. This is the way I will follow now to arrive in the end to what can be regarded as the deepening and, at the same time, extension of the scope of the dialectic-holistic understanding of human life of the mind-spirit by way of analyses of autobiographies that Dilthey and Misch considered to be the most authentic expressions of it.

One of the reasons for this choice is that already Dilthey himself seems to have had the intention to go in this direction according to the remnants of his attempts to develop the critique of historical reason. The following sentence is one of several witnesses. “Let us consider autobiographies, which are the most direct expression of reflection (*Besinnung*) about life” (Dilthey 2002. 219).

A superficial reading of this sentence is enough to let the reader suspect that no randomly chosen autobiographies will meet this high standard which can be reversed and transformed to become a requirement: real autobiographies can only be considered those that are the most direct expression of reflection (*Besinnung*) about life. Indeed, Dilthey enumerates the most eloquent and most elaborated autobiographies in European literature: Augustine, Rousseau, and Goethe are his paradigmatic examples. Characteristically, they will come to the fore again when we will investigate how Misch consummated Dilthey’s commencement.

To render the first superficial reading more profound we have to show up how Dilthey introduces in the context of autobiographies those categories of the “critique of historical reason” that are particularly apt to grasp what takes place when the understanding of finite human life of mind-spirit appropriately experiences, conceives itself in order to formulate in *autobiographical* narratives its findings and constructions. Viewed from the project of the critique of histor-

ical reason as part of a theory of knowledge the commencements of which are outlined in the first part of this paper, what takes place in this case is radically different from what happens when natural scientists report on the results of experiments in their labs or other artificial environments. Natural scientists are absolutely not interested in constructing their complex mental-cognitive operations from those elementary ones onward the description of which constitutes one of the main target of the Diltheyan philosopher. Nor are they attracted by the task to reflect upon the conditions of possibility of having “objects” at their disposal. By contrast, the Diltheyan philosopher turns back the direction of the cognitive attention from the “given” external objects to be grasped to the extremely complex operations of the more or less conscious mind that make it possible first of all to grasp objects conceptually. Even if it can be seen as a somewhat polarised picture, nevertheless, one can tentatively maintain that in natural sciences, there are clear-cut roles: the impartial spectator follows the intentionally triggered interactions between well-defined entities that are basically separated from herself, their environment, and each other. Consequently, the appropriate style of account will be the shortest possible report on those causal influences that she as scientist observed in this artificial situation.

Contrary to this, the entities which the life of mind-spirit is bothered with are much less well-defined, separated ones influencing each-other as the impartial spectator deliberately provokes them to do. The spectator is also far from being impartial because she reflects upon herself *as* reflecting upon herself and her environment. Thus the entities to be observed are incessantly shaping themselves and the “spectator” herself according to the dialectical-hermeneutical relationship of the parts-whole circles. The main categories Dilthey advises us to use when conceiving these processes are *meaning, value, sense, and purpose*.

In this respect, there is a certain ambiguity in Dilthey’s text that is mirrored in my own above formulation about the individual who narrates its *findings and constructions* respectively. Because in order to find something, it must be *given in advance*, whereas *construction* is precisely needed where what we are looking for is *not pre-given*. The ambiguity in Dilthey’s text lies in his answer to the question concerning the activity and passivity in our relationship with what he terms “meaning” or “sense” (*Bedeutung, Sinn*). On the one hand, in the closing passage of his section that has the title “The life-nexus”, he employs an almost “dogmatic” formulation.

Each life has its own sense. It consists in a meaning-context in which every remembered present possesses an intrinsic value, and yet, through the nexus of memory, it is also related to the sense of the whole. This sense of individual human existence is unique and cannot be fathomed by conceptual cognition; yet, in its way, like a Leibnizian monad, it represents the historical universe. (Dilthey 2002. 221.)

L. Tengelyi's strict refusal of the idea of "life as it writes itself"<sup>2</sup> must have originated in this or similar passages that are as many witnesses of the lasting influence of Leibniz's interpretation of the traditional concept of a monad. And to highlight the result of the promised more profound reading of the above sentence, it is more than evident that only those autobiographies can be regarded as paradigmatic that were written by individuals whose lives "represent the historical universe" extensively and intensively at the same time. This parallel expectation of a monadic individual and its adequate autobiography became the absolute point of orientation for Dilthey and Misch when fashioning their idea of the *philosophically analysable* individual and autobiography: the appropriate subjects were distinguished authors of world-literature, world-historical individuals of their own rights. When analysing Misch's work, we will see how this point of orientation shifts gradually and receives a historically more extended sense – without, however, giving up the basic parallel expectation itself.

The other tendency of the ambiguity in Dilthey appears in passages in which he stresses the activity of the individual in question when instead of passively accepting senses of longer life-units or the whole life she actively shapes and reshapes them again and again not presupposing a pre-given monadic eternal sense or meaning. This tendency promises to be more adequate for a "post-modern" understanding of individuality, identity, and sense or meaning of life – including perhaps that of L. Tengelyi's understanding. What the English translation terms "reflection" is *Besinnung* in the original. This can certainly be rendered by "reflection". Yet, the original word can well be understood as referring to an active attitude toward the "sense" (*Sinn*), which could be rendered by "providing with a sense", similarly to the German words *Bedeutung*, *Gestaltung*, *Entwicklung* in contexts such as the one referred to in our following quotation – immediately preceding the former one:

The sense of life lies in *giving shape to things* (in German simply: *in der Gestaltung*) and in development (*in der Entwicklung*); on its basis, the meaning (*Bedeutung*) of the moments of life is determined (*bestimmt sich*) in a distinctive way; it is both the experienced, intrinsic value (*erlebter Eigenwert*) of the moment and its productive force (*wirkende Kraft*) (Dilthey 2002. 221).

In this quotation, the use of the expression "productive force" of the moments of life – Dilthey's term is *wirkende Kraft* – suggests another analogy to Leibniz who deliberately employs the usually physical expression "force" in the context of the phenomena of the mind, and he attributes to the force inherent in the mental a more distinctive status than to the physical force usually taken to be the original.

<sup>2</sup> Cf. Tengelyi 1998. 24–26.



That the ambiguity concerning the respective roles that life itself and the individual author of an autobiography are supposed to play is a real one can be seen in Section 4 of the *Project* that is titled straightforwardly “Autobiography”. Dilthey explains the situation as follows:

Thus the initial tasks involved in apprehending and explicating a historical nexus are already *half solved by life itself*. [...] A coherence is formed within life itself, albeit from different standpoints and with constant shifts. The work of historical narrative is *already half done by life itself*. (Dilthey 2002. 221, 222.)

Dilthey relies on life when maintaining life’s own half-works in autobiographies to provide the human historical sciences with the basis of objectivity he believes they are in need of. He does not seem to be bothered particularly with what happens in the respective other halves provided by the individual authors. A pre-Freudian thinker as he was, Dilthey does not seem to have doubts that a rich, profound and extensive autobiography of a self-conscious historical individual warrants both the paradigmatic greatness of the author’s individuality and that she is the lady in her soul’s house, and so she is not influenced by unconscious forces of her soul to the effect that the truth of the content of the narrative is not to be called in question. Questions of sincerity or distortion, counterfeit, conscious or unconscious do not seem to appear on Dilthey’s horizon as disturbing factors to be taken seriously. He embraces as an imperative for all historical research aiming at objectivity that it has to start with the historical reports conveyed by the extant autobiographies of the great historical individuals.

This imperative is the real guiding principle for the life-long enterprise of Dilthey’s most talented pupil who happened to become his son-in-law: Georg Misch. Misch begun working on the history of autobiographies in the years when Dilthey composed his *Critique of Historical Reason*. One can see in Misch’s historical texts the unfolding of Dilthey’s basic *philosophical* idea as an archetype in the context of real *historical research*.<sup>3</sup>

The first concise manuscript version of his history of autobiography was conceived and written as a prize essay for the Prussian Academy of Sciences that appreciated his efforts and elected him as the winner of the competition. The first printed volume appeared shortly afterwards (1907) whereas the last one was published by his friends posthumously (1969), based on the original concise

<sup>3</sup> After we have learned the strong philosophical presuppositions behind the parallel expectations of a real autobiography and a great historical individual, there are some reasons to suspect that the presuppositions guide the enterprise also in the sense of prohibiting the acknowledgment of some texts as autobiographies and vice versa. This is a complex issue the treatment of which will require a separate investigation. Some hints are given in the present paper as well when the shift in Misch’s perspective on history is mentioned.



version – i.e. unfinished if compared to the extensively reworked texts of the earlier volumes.

There can be no reasonable doubt concerning Misch's complete agreement with Dilthey's view on autobiography including the concept of *life describing itself* – at least halfway, as we have seen. He intended to transplant Dilthey's above mentioned general imperative into the body of descriptive research by way of collecting and describing with great precision each and every autobiography from ancient antiquity through the 18<sup>th</sup> century *as* the foundation of all historical research. Misch shared Dilthey's reliance on life as a quasi-author collaborating with those great individuals whose greatness is certified precisely by their complete and well-articulated autobiographies. So much so that he maintained as late as ten years after World War 2 the legitimacy of the claim to autobiographies' providing us with a distinctive access to objective historical truth. In the series of radio-talks called "*Funk-Universität*" in the framework of the RIAS (*Rundfunk im amerikanischen Sektor*), in a broadcasting on 5<sup>th</sup> and 6<sup>th</sup> August 1954, he talked about *The Problem of Truth in the Autobiography*.<sup>4</sup> Already the first sentence of this talk evidently shows how completely he shares Dilthey's view: "If we pose the question of truth, the autobiography must evidently be preferred to the biography."<sup>5</sup>

The second sentence also argues for this view in a way that makes evident the Diltheyan roots of the statement. "For if there is a will it seems easier to achieve self-knowledge than to make a just judgement about another person, or at least self-knowledge can penetrate deeper."<sup>6</sup>

At the same time, however, it displays involuntarily the inadequacy of this very view with respect to such obvious historical facts as that the radio sender was not far from the Citadel in Spandau in which the imprisoned Albert Speer was about to complete his autobiographies with the titles *Inside the Third Reich* and *Spandau: The Secret Diaries* whose truthfulness has been strongly disputed, to say the least.

Yet, Misch is not as naïve as ignoring the authors' virtually ineradicable tendency to deception; but he believes this is only relevant in the political autobiographies.

<sup>4</sup> The typescript is kept in the University Library of Freie Universität Berlin (ZVN 104). It bears the title *Das Problem der Wahrheit in der Biographie*. This must, however, be a typing mistake; already the first sentence is unmistakably about the *autobiography*. The typewriter must have mistaken this broadcast with the next one in the series which is indeed about the biography: "Das Problem der Wahrheit in der Biographie" von J. Romein.

<sup>5</sup> "Wenn die Wahrheitsfrage gestellt wird, hat die Autobiographie augenscheinlich einen großen Vorzug vor der Biographie."

<sup>6</sup> Denn Selbsterkenntnis ist, so scheint es, wenn der Wille dazu da ist, leichter erreichbar als die gerechte Beurteilung eines anderen und vermag jedenfalls tiefer zu dringen.

For political autobiographies and even more the courtly-political memoirs are usually murky sources; they do not originate in the will to truth but in the need of self-justification. Contrary to this, the poets and some philosophers have predilection for autobiographical writing, and this is well-founded.<sup>7</sup>

In the main part of the short text, Misch wishes to elucidate the sense in which Augustin's *Confessions* can be understood as an example of the truth of an autobiography. First he informs the audience based upon convincing philological evidences that Augustin described his conversion immediately afterwards as a result of Platonic influence much more than the grace-based influence of the Gospel's and St. Paul's teaching. In spite of this fact, however, he insists on the truth of the Church Father's mature story that unambiguously reverses the order of influences:

This procedure does not happen by chance but it corresponds to a general law of development of higher order autobiographies. For the development of such autobiographies, it is decisive that their authors have understood their results in their life-experience. How should the autobiography be able to testify something else than the awareness of the individuals of themselves, which they have accomplished by that time? But it is precisely by virtue of the springing from the present life-understanding of the authors binding together the past that the autobiography gains the force to shape and reshape the historical facts that lies in the sense of the experienced life. This force by virtue of which they are capable of elevate themselves to the level of poetry, and to arrive to the norm that we usually term through Goethe's double concept of 'Truth and Poetry' – without ceasing to be bound by historical evidence.<sup>8</sup>

In fact, it is tempting to understand the last sentence as implying a deeply felt identification with the history of German spirit (*Geistesgeschichte*) that could be made explicit by way of completing it as "wir [Deutschen]" instead of having

<sup>7</sup> "Denn politische Autobiographien und gar höfisch-politische Memoiren sind in der Regel recht trübe Quellen, da sie nicht aus dem Wahrheitswillen sondern aus dem Bedürfnis der Selbstrechtfertigung hervorzugehen pflegen. Dagegen haben die Dichter und auch manche Philosophen eine Vorliebe für das autobiographische Schrifttum, die gleichfalls wohlbegründet ist."

<sup>8</sup> "Dieser Fortgang ist nicht Zufälliges, sondern entspricht einem allgemeinen Bildungsgesetz der Autobiographie höherer Art, demzufolge für ihre Ausbildung ein übergreifendes, von den Autobiographen in ihrer Lebenserfahrung errungenes Verständnis ihrer Ergebnisse maßgebend ist. Wie sollte die Autobiographie auch von etwas anderem zeugen können als von dem jeweils vollendetem Bewußtsein der Individuen über sich selbst? Aber gerade durch dieses Quellen aus dem gegenwärtigen Lebensverständnis des Autors, das die Vergangenheit in sich zusammenhält, gewinnt die Autobiographie die Kraft zu einer Gestaltung und Umgestaltung der historischen Fakta, die im Sinne des durchlebten Lebens liegt: die Kraft, mit der sie ohne von der für die historischen Werke unaufgebbaren Bindung an den Stoff zu lassen, sich zu dem Niveau der Dichtung zu erheben und also die Norm zu erreichen vermag, die wir [fast hört man mit: »wir Deutschen«...] durch den Goetheschen Doppelbegriff 'Wahrheit und Dichtung' zu bezeichnen gewohnt sind."

simply “wir”. Who else’s association could be meant by the first person plural when stating that “we usually term [it] through Goethe’s [...] ‘Truth and Poetry’”? This presupposition leads to the deeply problematic inner life-narrative of Misch himself, which has not at all been atypical for Jewish intellectuals with a profound German cultural-national identity who survived the period of the Nazi regime and the Holocaust. Howsoever it may be, there is a philosophically interesting parallel between the facts that *against all that had happened* in the tragic catastrophes of the 20<sup>th</sup> century Misch insisted on a – *sit venia verbo* – Apollonian philosophical understanding of autobiography that was an understanding deeply rooted in German cultural and spiritual history. Also in this sense, he followed Dilthey who wrote a series of important works on German cultural-spiritual history on the one hand, and, on the other, declared his view about the special categories that play a role in autobiographies as follows: “The category of purpose, *or of the good*, which considers life as it is directed toward the future, presupposes that of value.” (Dilthey 2002. 223, emphasis added)

Even if negative values also can occur in autobiographies, the *par excellence* category of value meant by Dilthey is the positive, the good. According to the well-known Aristotelian schema, the purpose is interchangeable with the good. But the question can hardly be avoided, how this supremacy of the good could remain intact even in an age for which Walter Benjamin maintained that novels such as Döblin’s *Berlin, Alexanderplatz* testified to the disappearance of distinguished positive values attached to positive heroes representing a nation or their own education with the aim of acquiring and embodying those values through their virtuous behaviour. Not to mention the gods of Brecht vainly looking for a good person in Szechwan...

In this context, it can be instructive to recall that Misch appropriated Dilthey’s view implying the concept of autobiography as the *life describing itself halfway* together with his *reliance on life* as a quasi-author collaborating with great individuals, the authors of those autobiographies. Their greatness is warranted precisely by their complete and well-articulated autobiographies that mirror those human values which can be found at least explained in the most profound works of German literary spirit, autobiographical or fictional-educational. My reason for recalling this now is to emphasise that not only did Misch suffer from a particular type of split-mindedness as a German-Jewish intellectual but also as a philosopher and a historian at the same time and in the same mind. For he was a philosopher of Dilthey’s denomination concerning the distinctive treatment of the great personalities. He goes as far as constructing the whole edifice of his history of autobiography on the thesis of the gradual development of the self-conscious individual. The introduction of the first volume to the *History of Autobiography* allows no doubt as for this principle of construction:

Biography [...] became a definite literary form among the Greeks [...] its basis was this same conception of *bios*. [...] It was essentially concerned with the unchanging ego of the fully matured human being [...] This original reality of a man's character, which in truth gives consistency to his life, should not, however, be separated from the shape life takes [...] We meet here with the essential connection of the ethical with the aesthetic that is expressed in the term *kalokagathon*. This [...] classical Greek conception of personality was at work not only in the formation of biography but also in other literary forms available for the description of individual life [...] (Misch 1950. 62–63.)

Although Misch clearly recognizes that this – Apollonian – concept of an ideal human being *was* an abstraction, separated from the way the individual human beings *became* necessarily imperfect embodiments of the ideal, the principle of his own work – following Dilthey and Burckhardt – originated in the late reflex of this same concept:

[...] some Renaissance writers mastered one of the crucial problems of biography – that of seeing at the same time [...] the broad conception of an individual, and [...] the characteristics of which it is built up. And it is in the field of autobiography that this achievement is found. [...] Indeed, in face of the formless flow of the narrative of later books of confessions, it must be said that the classical attitude, demanding from the autobiographer both form and style, has permanent validity. Was that not Goethe's own rule? [...] [The causes] are contained in the very general obedience to historic law which has determined the development of Western autobiography. [...] For autobiography there is laid down by the very nature of its subject matter a law of development [...] (Misch 1950. 64–65.)

[...] in ancient European literature autobiography [originated] through the individual's reflective awareness of his personality. [...] however late, it became the organ for the expression of individuality. If we are to build our history on firm foundation, it must proceed from the beginnings of consciousness of personality, which was a present from the Greeks to the European world. (Misch 1950. 69.)

So the firm law of autobiography is that it followed the development of the self-awareness of individual personality. But what happens, if or when this self-conscious personality disappears? In such a situation, the autobiography will either disappear itself or have its meaning profoundly transformed. The shared conviction of Dilthey and Misch was that precisely this happened after Goethe.

One of the reasons, however, why I have mentioned the split-mindedness of Misch can be summarised briefly without going into much details: in later volumes of his *History of Autobiography* he seems to have immersed in professional historical-philological research, and to some extent, this let him dissociate

himself from the preconceptions of Dilthey's and Burkhardt's. Today historians of ideas with much micro-philological precision specialised in the respective historical periods that Misch investigated appreciate unanimously his methods and works *as that of an outstanding historian*.

The other reason to mention the split-mindedness is philosophical. For more than one prominent representatives of philosophy in the 20<sup>th</sup> century invested much effort into analyses of the philosophical, anthropological, artistic, and other implications of the disappearance of the self-conscious Goethe-type individual having basically the good as its purpose. One of them was Heidegger especially in *Being and Time*. And Misch was aware of both the importance and the shortcomings of Heidegger's work from the perspective of the original Dilthey-an philosophical motives. But he must have been aware of the shortcomings of the original Diltheyan philosophical perspective at least indirectly by way of his historical findings without being able to transform it in a way that would have issued in an elimination of the shortcomings of Heidegger's work replacing it with another synthesis of Husserl's and Dilthey's respective thought that would have done justice to both master thinkers.

The last issue to be tackled in this paper is a historical-methodological one that I would only mention briefly.

Our earlier quotations already insinuated the shared conviction of Dilthey and Misch that the series of autobiographies to be taken seriously from a philosophical-historical point of view came to an end with Goethe's *Dichtung und Wahrheit*. They were obviously convinced that this is the greatest ever autobiography written by the greatest ever personality who fully obeyed the historic law of autobiography.

Misch was aware, however, of the fact that at least works that belonged to the formal genre of autobiographies were also written – and abundantly, for that matter – in the late 19<sup>th</sup> and early 20<sup>th</sup> century. Facing this lively interest in this genre, how could he react if unable to follow either Heidegger's or Plessner's way? He believed that those writings, a side-shoot of the original genre that called for *philosophical* interpretations can only awaken some *sociological* interest. He must have been convinced that the authentic personality creates a narrative framework within which her innermost individuality confers meaning to the past, authenticates her presently felt values and projects her good purposes onto the future. This is the categorical structure of *her* unified consciously told narrative that unanimously supports her claim to authenticity in a deep historical and philosophical sense.

Contrary to this, the new type of human personality is characterised by having a number of sociologically defined roles to play, and she is, at best, only struggling for an identical narrative warranting her stable personal identity – far from even posing the question of authenticity.

As opposed to the Dilthey-Misch type of historians supported by their philosophies of life, today historians and literary scientists do not reckon with deeds

and events narrated in higher order autobiographies of unique individuals as the chief sources of historical knowledge. If they choose an at least quasi-substantial factor to have a mooring within the flux of historical life it is in no way the great personality as such whom they elect to render coherent and meaningful the aggregate of the disparate facts. It is either the working of impersonal structures or the appearance of the similarly impersonal events (*Ereignis* – what strikes the eye). They rely on micro-historical or contract-theoretical (Lejeune) reconstructions of the ways of life of people whose lives were basically determined by playing their appropriate roles without any authentic decisions from their innermost essence about meaningfulness, values, or purposes. Also the methodological use of *long durée* processes have come to the fore. As a consequence of these changes, the new forms that the main methodological problem in the historical sciences put on, and that were circumvented by renowned scholars in volume 5 of *Poetik und Hermeneutik* – (*Geschichte – Ereignis und Erzählung*; cf. the studies in Koselleck 1973.) is as follows: what are the roles like that the classical (world-) historical individual can be attributed in the process of constructing history – *res enarratae* – as history of *salient events*. “Event” is meant here a particular event elevated from the flow of life and made almost a closed whole by a *personal* actor – this is the equivalent of the German word *Ereignis*. Or how did – if at all – “diachronic structures” develop from the socialising-socialised individuals or the global structures far below the scale of *personal* human lives; how can/must they be regarded as the main objects of a historical representation. Some historians distinguish two genre to be used when writing historical texts. They maintain that the former is connected to the genre of *narration* (*Erzählung*) whereas the latter is attached to that of *description* (*Beschreibung*). One of the leading scholars of that generation, historian Reinhard Koselleck considers them – structure and event, description and narration – not excluding but conditioning, completing each other, in a way that can, in fact, easily be transformed into a hermeneutic circle, and thus connected to the methodological efforts of Dilthey and Mich.

## REFERENCES

- Boros, Gábor 2017. Személyes azonosság, érzelmek, elbeszélés (narrativitás). In Bartha Judit – Gyenge Zoltán (szerk.) *Töredékes dialektika. Írások Weiss János 60. születésnapjára*. Budapest, L’Harmattan. 59–68.
- Dilthey, Wilhelm 2002. *The Formation of the historical world in the human sciences*. Princeton/NJ, Princeton University Press.
- Goldie, Peter 2014. *Me Inside. Narrative, Emotion, and the Mind*. Oxford, Oxford University Press.
- Hobbes, Thomas 1966. *Thomae Hobbes Malmesburiensis Opera philosophica quae Latine scripsit omnia*. Vol. 1. Aalen, Scientia Verlag.

- Koselleck, Reinhart – Wolf-Dieter Stempel (eds.) 1973. *Geschichte – Ereignis und Erzählung*. München, Fink Verlag.
- König, Josef 1937. *Sein und Denken. Studien im Grenzgebiet von Logik, Ontologie und Sprachphilosophie*. Halle, Niemeyer.
- Misch, Georg 1931/1967. *Lebensphilosophie und Phänomenologie. Eine Auseinandersetzung der Diltheyschen Richtung mit Heidegger und Husserl*. Stuttgart, B. G. Teubner.
- Misch, Georg 1950. *A History of Autobiography in Antiquity*. Part One. London, Routledge and Kegan Paul.
- Misch, Georg 1994. *Der Aufbau der Logik auf dem Boden der Philosophie des Lebens*. Freiburg (Breisgau) – München, Alber.
- Tengelyi László 1998. *Élettörténet és sorsesemény*. Budapest, Atlantisz.

TIM THORNTON

## Narrative Identity and Dementia

### I. CO-CONSTRUCTIONIST CLAIMS FOR NARRATIVE IDENTITY IN DEMENTIA

It seems obvious that one of the harms that dementia does, both directly to the person who develops it and indirectly to their kith and kin, is to undermine the person's identity. One reason for thinking this is that, since John Locke's discussion of it, personal identity has been associated with continuity of a subjective perspective on the world held together by memory and that memory is severely curtailed in dementia. Hence dementia seems to threaten an individual's identity as a particular person, gradually undermining it.

But the necessity, or the closeness, of the connection has been criticised by a number of philosophers and healthcare professionals who subscribe to a narrative account of personal identity. Their argument goes as follows. If personal identity is constituted through a personal narrative rather than, for example, a memory connection, then while the capacity to author a self-narrative also seems to be threatened by dementia, that need not undermine personal identity providing that the narrative that constitutes identity can be *co-constructed*. As dementia takes hold, authorial responsibility can fall to others.

Clive Baldwin, a professor of narrative studies with a social work background, argues in this way. First, he claims that human subjects have narratively constituted selves and hence, pessimistically, are susceptible to harm via that narrative in, for example, the case of dementia.

[W]e are indeed narrative beings who find our Selves in the stories we tell about ourselves and the stories that others tell about us; that narrativity is essentially an inter-personal activity; that some people find their stories marginalized, themselves as narrators dispossessed; but that it does not have to be that way. The stories we tell... can subvert the status quo and open the door to new ways of telling, and thus new ways of being. I will develop this argument through the lens of the experience of people with dementia, though it has been argued elsewhere that people experiencing severe mental illness may also be narratively dispossessed. (Baldwin 2008. 223.)



But, more positively, the threat of such harm can be turned aside through the joint authorship and co-construction of identity-constituting narratives.

[W]e look towards the joint authorship of narratives where the narrative process is shared by people living with dementia and those around them. This may take the form of co-construction of narratives (see Keady & Williams 2005) whereby the final narrative is very deliberately and consciously a negotiated product between those people living with dementia and others or the piecing together and progression of the fragmented narratives of people living with dementia by those who support them. (Baldwin 2008. 225.)

The philosopher of psychiatry Jennifer Radden and psychotherapist Joan Fordyce deploy a similar argumentative strategy. First, they subscribe to a form of narrative identity theory concerning what Marya Schechtman calls ‘characterization identity’: “the set of characteristics each person has that make her the person she is” (Schechtman 1996. 74).

A person’s identity comes in the form of a self-narrative in the work of many who employ these categories. [...] The actions and experiences making up that narrative comprise the personal story of which the subject stands as ‘author’. (Radden and Fordyce 2006. 73.)

Such self narratives are always, they suggest, co-constructed, though generally this is tacit. But in the case of dementia, the relative contributions to authorship change and become more noticeable.

The construction and sustaining of the person’s characterization identity have been, until the deficits of dementia make themselves known, collective efforts conducted largely tacitly. Increasingly, as these deficits erode aspects of the person’s memory and self-awareness, the task will come to include the provision of explicit identity recognition – a response that says, in some form, ‘this is who you are and what you are like’ [...] Until now, also, to the extent that others were called on to sustain the identities of those around them, this task will have been largely mutual. Other people will have helped sustain, just as they helped constitute, my identity at the same time as I helped maintain (and constitute) theirs. Now, however, the task of holding and preserving the identity of the person suffering dementia will come to be placed more squarely on the shoulders of others (often, these are the shoulders of second persons, intimates, and the customary societal carers, women). (Radden and Fordyce 2006. 81.)

It might seem that this account is *too* optimistic. If personal identity is constituted by self-narratives that can be co-authored then providing that caregivers

or kith and kin are ready to step up to the breach then dementia is no longer a threat to identity. Radden and Fordyce concede that this is not how it *seems*, however.

The most noticeable initial problem with this model is perhaps the discomfort and sense of falsity it sometimes brings upon those others left with the burden of sustaining the identity of a loved one through these processes of holding, reinforcing, and reinscribing. Although perhaps a distorted reaction, the response is often angry and disappointed. The loved identity seems to have gone – replaced by an alien changeling, it sometimes seems, or by no one. ‘This is what you were and were like’ we want to say to the dementia sufferer, ‘but no more!’ [...] The heart-breaking aspect of this task of sustaining characterization identity cannot be ignored. Nonetheless, it is an enterprise apparently required by the very notion of characterization identity as that identity has been defined and explained here. (Radden and Fordyce 2006. 81–82.)

Taking the relevant sense of personal identity to be Schechtman’s characterization identity and taking that to be constituted by a co-constructed personal narrative, it follows, they claim, that there is a normative requirement on carers however angry and disappointed they may feel at the misleading appearance of the loss of their loved one. They do not, however, explain the nature of this obligation.

There is a more significant problem with this idea which can be illustrated by an example from the other end of life: it would allow the sincere ascription of youthful authorship of the “round robin” letters sometimes written in the UK “as from” small babies around Christmas. Such ascription would simply require a generous interpretation of a baby’s still limited behavioural repertoire by dotting parents through which the meaning and thus authorial intention would be constructed, rather than revealed. There would be no further issue of whether this accurately tracked antecedent communicative intentions. Whilst in the case of such round robin letters no abuse – except perhaps of good taste – is risked, in the case of dementia the construction of a narrative by only one party in a supposed conversation does carry that risk.

Stephen Sabat, who has done much to promote the idea that even advanced Alzheimer’s sufferers may still be ‘semiotic subjects’, gives one such example:

In many cases, caregivers often do attribute intention to the afflicted person in that caregivers may believe that he or she is acting deliberately to annoy them, when in fact the annoying behaviour is due to cognitive impairment. If the afflicted person’s recall memory is severely affected, he or she may ask the same question repeatedly. This is hardly due to an intention to annoy anyone. It is of utmost import that caregivers identify the circumstances in which intention is present and healthy and not meant to annoy. (Sabat 2001. 222.)

The idea of co-construction is particularly dangerous in psychiatry because of its history of paternalism. Humane responses to that history have stressed the perspective of individuals, the importance of respect for autonomy and patient values even where these are hard to discern. Suggesting that personal narratives, and hence selves, can be *made up* by others seems a complete abandonment of the rejection of paternalism by the most insidious of means. So why has the idea of constructing those, supposedly on someone else's behalf, come to seem a humane response to dementia? I will argue that it follows from misrecognising the fundamental difference between this dangerous, paternalistic invocation of co-construction of personal identity and the innocent role of constructionism in response to an issue that looks superficially similar: asking whether someone is still the same person as they were before dementia but where the word 'same' is used in Wittgensteinian secondary sense. What may look merely like a subtle difference makes all the difference.

The structure of this paper is, sadly, quite complex. Starting, here, from Radden and Fordyce's unfortunately paternalist account of identity and dementia, I will work 'backwards' and then 'forwards'.

Radden and Fordyce's account is based on Schechtman's 1996 narrative account of personal identity. Schechtman argues for her narrative account by saying that it is a good answer to what she calls the 'characterization question', which she contrasts with the 'reidentification question'. She rejects the reidentification question because neo-Lockean attempts to answer it fail. In arguing for this, however, she ignores the best neo-Lockean approach: McDowell's anti-reductionist version. This is a defect in her argument given that it is in part, at least, an argument from elimination.

Working "forwards", Schechtman's answer to the characterization question is a substantive narrative theory of identity but both that question and her answer is ambiguous between a notion of basic identity and a richer notion of a moral subject. Because she says that she builds her account from two others, which I will characterise – following her citations – as Dennett's and MacIntyre's, I use these to assess the two interpretations or aspects of her account. There is, however, independent reason to reject Dennett's account – and anything like it – leaving MacIntyre as the only plausible model of a narrative account. But his account does not support Radden and Fordyce's stronger claims about co-construction. Further, since Schechtman's account is motivated by an *argument from elimination* that ignores a better option, it is not clear we need a substantive account anyway. Freed from that, a better way of learning lessons from MacIntyre to apply in the case of dementia is available. Narrative can help shed light on very specific identity questions asked in the case of dementia but in a different way to Radden and Fordyce's paternalism.

## II. MARYA SCHECHTMAN'S REJECTION OF THE REIDENTIFICATION QUESTION

Radden and Fordyce's account of co-construction of identity is based on their modification of Marya Schechtman's narrative constitution view of personal identity. That in turn is her proposed answer to what she calls the "characterization question" which she contrasts with the more familiar reidentification question in the philosophy of personal identity.

Most simply put, this [characterization] question asks which actions, experiences, beliefs, values, desires, character traits, and so on... are to be attributed to a given person. Reidentification theorists ask [by contrast] what it means to say that a person at  $t_2$  is the same person as a person at  $t_1$ ; characterization theorists ask what it means to say that a particular characteristic is that of a given person. (Schechtman 1996. 73.)

Schechtman prefers the characterization to the reidentification question and her proposed account of identity is an answer to the former rather than the latter. It might thus seem that, by answering a distinct question, it is incommensurable with answers to the latter question proposed by other philosophers, especially those working in a broadly Lockean framework. But although there is one relevant difference (to which I will return), I think that Schechtman takes her narrative constitution view to be an account of personal identity, however precisely that is to be understood, and hence to be a competitor to neo-Lockean accounts. Before returning to her answer to the characterization question, I will briefly sketch the nature of the reidentification question.

In a more recent book, Schechtman summarises her earlier approach thus:

I thus suggested that we instead think of the problem of personal identity as one of *characterization* – the question of which actions, experiences, and traits are rightly attributable to a person. The answer to a question of personal identity can then take the form of a relation between persons and psychological elements or actions *rather than of a relation between time-slices*. (Schechtman 2014. 100, italics added.)

The contrast with a relation of time slices stems from a view of personal identity that derives from an interpretation of John Locke who said:

To find wherein personal identity consists, we must consider what person stands for; which, I think, is a thinking intelligent being that has reason and reflection and can consider itself as itself, the same thinking thing in different times and places. (Locke 1975. II. xxvii. 9.)

Person, as I take it, is the name for this self. Where-ever a Man finds, what he calls himself, there I think another may say is the same Person. It is a Forensick Term appropriating Actions and their Merit; and so belongs only to intelligent Agents capable of a Law, and Happiness and Misery. (Locke 1975. II. xxvii. 26.)

Locke thus suggests that personal identity has, and depends on, a continuity of inner perspective. To illustrate this, he considers a case in which the “Soul of a Prince, carrying with it the consciousness of the Prince’s past Life, enter[s] and inform[s] the Body of a Cobbler as soon as deserted by his own Soul”. In such a case, he claims that “every one sees, he would be the same Person with the Prince, accountable only for the Prince’s Actions” (Locke 1975. II. xxvii. 15).

This has inspired a philosophical industry concerning the idea that being the very same person, in the forensic sense of the person who should be punished for the earlier self’s crimes, is constituted by a kind of internal consciousness of identity over time. And then, so the thought goes, if that is the case, it ought to be possible to give an account of this continuous inner awareness in terms which do not presuppose sameness of the person over time because the aim is to define the latter using the former.

There are, then, some familiar questions. Is it really the case that events that someone does not recall cannot be part of their temporally extended existence as a subject? And does not memory *presuppose* the identity of the self/person because memory is awareness of things that have happened to oneself, not just historical knowledge in an impersonal manner? Various solutions have been outlined.

Schechtman argues, however, that none of the standard answers to the re-identification question in the Lockean tradition are successful. For that reason, she recommends swapping questions and then proposing a narrative answer to her preferred question. Before considering that, it is, however, worth highlighting an option she ignores.

As John McDowell argues, there is no need to assume that a reductionist project must work (McDowell 1994. 325–340). He suggests instead that we should *not* take Locke to be trying to reduce personal identity to continuity of inner awareness (and failing at that because Locke says explicitly: “can consider itself as itself, the same thinking thing in different times and places” which presupposes sameness). But, rather, Locke is pointing out non-reductively that it is a feature of persons that they have an inner perspective on their lives which gathers together events as theirs without any criterion or test of identity and without even the exercise of a skill in picking themselves out (contrasting the way that one might keep track of one of the red balls in a game of snooker). One does not have to identify oneself to oneself, one’s memories as one’s own rather than someone else’s. But that is not because one is a special locus of “mind-stuff” as Descartes assumed. No, it is because one is a body, with bodily criteria of identi-

ty, but one which happens to have – as a human – an inner perspective too which goes hand in hand, effortlessly, and, in general, agrees with those bodily criteria.

What are those bodily criteria for the identity of persons over time? Here, McDowell rather breezily suggests spatio-temporal continuity *under a sortal*. One way to make this clear is to imagine an alien with a very different kind of bodily life – perhaps as a cloud of gas – studying plant and animal life on earth down as far as the cellular level. As a rabbit, for example, lives, it eats grass and excretes dung. Thus vegetable matter gets merged with the rabbit and separated. Over time, there are complex chains of connection. But the spatio-temporal continuity of any particular rabbit does not have to take account of the grass and the dung with which it is brutally continuous: but rather the career of the rabbit itself rather than its food or dung. In other words, an appeal to spatio-temporal continuity is not a reductionist explanation of rabbit identity over time. Rather, the relevant mode of spatio-temporal continuity presupposes the sortal *rabbit*. The same applies to persons though with some complications.

One such complication is raised by the science fiction cases of the sort Locke himself considers: the mind of the prince transported into the body of the cobbler. In such a case, identity goes with the inner dimension rather than the outer body. But that is not to say that, in general, we have an understanding of the inner dimension independently of, or more fundamental than, the normal bodily criteria of identity.

The fact that Schechtman ignores this possibility is one reason to be suspicious of her argument from elimination in favour of the characterisation over the reidentification question. There may simply be no need to articulate a narrative theory of identity in the first place. Putting this point on hold for the moment, I will outline the development of her answer to her favoured question.

### III. SCHECHTMAN'S CHARACTERIZATION QUESTION AND NARRATIVE CONSTITUTION ANSWER

In more recent work, Schechtman provides the following overview of her position.

I thus suggested that we instead think of the problem of personal identity as one of *characterization* – the question of which actions, experiences, and traits are rightly attributable to a person. The answer to a question of personal identity can then take the form of a relation between persons and psychological elements or actions rather than of a relation between time-slices. Such an account, I argued, will be non-reductive but still informative. In particular I urged that rather than thinking of identity-constituting psychological continuity in terms of overlapping chains of psychological connections properly caused, we should instead understand it in narrative terms, a revision

made possible by framing the question as one of characterization. We constitute ourselves as persons, on this view, by developing and operating with a (mostly implicit) autobiographical narrative which acts as the lens through which we experience the world. (Schechtman 2014. 100.)

The characterization question, and her answer to it, is, however, ambiguous. In asking which actions, experiences, and traits are rightly attributable to a person, it might be asking which are *authentic* expressions of the person, their *moral* selves, aspects of their *deeper* character by contrast with momentary whims or temptations, or the distortions of alcoholic high spirits, for example. Or it might mean simply which of all the actions in human history were those of a particular person. More prosaically, the latter might be asked by a detective of an act of theft: who did it?

Schechtman has conceded this point about her earlier 1996 work.

For many, the switch from the reidentification to the characterization question automatically signals a switch to questions about the moral self. There are some good reasons for thinking so – my first move in introducing the view is to draw a distinction between the “Who am I?” question raised by a confused adolescent (which I link to the characterization question) and the “Who am I?” question asked by an amnesia victim (which I link to the reidentification question). At the same time, however, I meant for the characterization question also to answer questions about attribution at the most fundamental level – not only which beliefs and desires are *truly* mine in the sense of the moral self, but which are mine in the most basic and literal sense. (Schechtman 2014. 102.)

Given this latent ambiguity in the question, how should her earlier 1996 answer – the account which informs Radden and Fordyce’s view – be assessed? Fortunately, in her earlier work, Schechtman suggests a clue. She reports that her narrative constitution view:

draws its inspiration from a number of sources both philosophical and psychological which argue either that persons are self creating [...] or that the lives of persons are narrative in form. Weaving strands from these discussions together with my own analysis, I develop a view according to which a person creates his identity by forming an autobiographical narrative – a story of his life. (Schechtman 1996. 93.)

I suggest that these two distinct strands are responses to distinct interpretations, or aspects, of the characterization questions and I will use two of the philosophers Schechtman cites as sources to examine these two strands: Dennett and MacIntyre.

### 1. *Persons are self-creating*

Schechtman cites Daniel Dennett as a philosopher who defends the idea that *persons are self-creating*. He claims that a self is a “centre of narrative gravity”. To outline his view, he suggests an analogy with the physical notion of a centre of gravity.

A centre of gravity is *just* an abstractum. It’s just a fictional object. But when I say it’s a fictional object, I do not mean to disparage it; it’s a wonderful fictional object, and it has a perfectly legitimate place within serious, sober, *echt* physical science. (Dennett 1992. 104.)

The idea of a centre of gravity is deployed within a branch of physics to describe and predict the behaviour of physical systems acting under physical forces. What a centre of gravity is depends on this theoretical context and it is one of the useful tools and ideas that go to make that context. The concept is one amongst others interdependent on a theoretical stance.

Selves are given similar treatment. Like centres of gravity or mental states, they are theoretical, even fictional, entities articulated within an interpretative theoretical stance.

A self is also an abstract object, a theorist’s fiction. The theory is not particle physics but what we might call a branch of people-physics; it is more soberly known as a phenomenology or hermeneutics, or soul-science (*Geisteswissenschaft*). The physicist does an *interpretation*, if you like, of the chair and its behaviour, and comes up with the theoretical abstraction of a centre of gravity, which is then very useful in characterising the behaviour of the chair in the future, under a wide variety of conditions. The hermeneuticist or phenomenologist – or anthropologist – sees some rather more complicated things moving about in the world – human beings and animals – and is faced with a similar problem of interpretation. It turns out to be theoretically perspicuous to organise the interpretation around a central abstraction: each person has a *self* (in addition to a centre of gravity). In fact we have to posit selves for *ourselves* as well. The theoretical problem of self-interpretation is at least as difficult and important as the problem of other-interpretation. (Dennett 1992. 104.)

What is the motivation for Dennett’s version of a narrative approach? I think that it is useful to consider the perceived alternative to it that Dennett rejects. He gives a clear statement of this in the following passage which starts with a brisk re-iteration of the advantages of his narrative account for describing psychopathology but also mentions the alternative to which it stands opposed.



We sometimes encounter psychological disorders, or surgically created disunities, where the only way to interpret or make sense of them is to posit in effect two centers of gravity, two selves. One isn't creating or discovering a little bit of ghost stuff in doing that. One is simply creating another abstraction. It is an abstraction one uses as part of a theoretical apparatus to understand, and predict, and make sense of, the behavior of some very complicated things. The fact that these abstract selves seem so robust and real is not surprising. They are much more complicated theoretical entities than a center of gravity. And remember that even a center of gravity has a fairly robust presence, once we start playing around with it. But no one has ever seen or ever will see a center of gravity. As David Hume noted, no one has ever seen a self, either. (Dennett 1992. 114.)

Dennett here supports his account by reinforcing the apparent robustness and reality of narrative selves on the basis of a comparison with the robustness of centres of gravity. Nevertheless they are "created" abstractions. And this contrasts with the other possibility: "creating or discovering a little bit of ghost stuff". I take it that more important alternative here is "*discovering* a little bit of ghost stuff" which stands in as a brisk summary of a Cartesian account.

It should come as no surprise that Dennett's main opponent is a form of Cartesianism, whether of a traditional immaterialist kind or a form of materialism which shares a key feature. That feature, and a key target of his *Consciousness Explained*, is the idea that "somewhere, conveniently hidden in the obscure 'centre' of the mind/brain, there is a Cartesian Theatre, a place where 'it all comes together' and consciousness happens" (Dennett 1993. 39). Even if Descartes' immaterialism is rejected, this idea can remain implicit in thinking about the brain.

Let's call the idea of such a locus in the brain *Cartesian Materialism*, since it's the view you arrive at when you discard Descartes's dualism but fail to discard the imagery of a central (but material) Theatre where 'it all comes together'. The pineal gland would be one candidate for such a Cartesian Theatre [...] (Dennett 1993. 107.)

I take it that an immaterial centre might constitute a self and a material centre could at least underpin one. Dennett rejects any such approach and deploys the narrative account as (part of) his alternative. This also explains his comment (above) that "[a]s David Hume noted, no one has ever seen a self". He continues by quoting with approval Hume's doomed attempt to spot his own self among his mental states.

For my part, when I enter most intimately into what I call *myself*, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch myself at any time without a perception, and never

can observe anything but the perception... If anyone, upon serious and unprejudiced reflection, thinks he has a different notion of himself, I must confess I can reason no longer with him. All I can allow him is, that he may be in the right as well as I, and that we are essentially different in this particular. He may, perhaps, perceive something simple and continued, which he calls himself; though I am certain there is no such principle in me. (Hume 1978. 252.)

Hume's final comment is clearly meant to be ironic. Introspection, Hume suggests, reveals nothing that could stand in the sort of relation to one's mental states that a self is supposed to do. This leads him to advocate a minimalist "bundle theory" of mind. The self is identified simply with the mental states encountered in introspection and not with an ego which stands in a relation to them.

Dennett shares Hume's opposition to a Cartesian ego but he adds a principle of organisation to the mental states gathered together via narrative. The self is not just a *bundle* of states but states *structured* in a narrative. Given his related account of mental states there is no tension between primitively real mental states and mere fictional selves. Both mental states and the narrative structure that adds up to a self are theoretical constructs. But it is also important to note that there is no antecedently understood author to the narrative. That idea would correspond to a substantial, pre-narrative self. Dennett's idea, by contrast, is that the self just is the structured narrative.

Given the choice between Cartesian "ghost stuff" and a narrative account, then the latter is obviously the more attractive. It also seems to receive support as descriptively accurate from both Hume's introspection and Dennett's "hetero-phenomenological" method which at least takes account of first person reports (whilst not uncritically according them apodictic certainty). But the choice is, nevertheless, a forced choice.

If Schechtman's narrative constitution account is viewed through the prism of Dennett's account, then it can be seen as answering the more austere question of basic identity, rather than to the richer notion of the nature of the moral self. Schechtman seems to take a narrative answer to the characterization question to be a rival to the reductionist version of the neo-Lockean account of basic identity. Sniffy about traditional accounts that argue that the different temporal stages of a person are unified by overlapping chains of memory, she raises a prior question: "what unites even at a given temporal stage the experiences that are those of a particular individual?" and answers narrative. Narrative unites at a given time. But it also unites over time because narratives are temporally extended wholes. Further, it suggests a nice distinction. On a narrative account, temporal elements are abstractions from a whole rather than free-standing elements needing uniting together. They are more like the pitch and timbre of a note than like the individual bricks that can be combined to make a wall.

Dennett, too, deploys narrative as a basic principle of organisation to unite actions, experiences, and traits in a single person. But although it is austere, it is nevertheless radical. Fully assessing such a radical claim about the connection between self and narrative lies outside the scope of this paper. But it is worth noting that, influential though it has been, it faces at least three significant difficulties:

First, what is the claim actually being made? Are selves narratives or are they the authors of narratives? They cannot be both. But in many published accounts, these two ideas are not consistently distinguished. Baldwin, for example, blurs both together because, in addition to the claim above that selves are stories, he also says:

If we are narrative beings and the primary narrative of our life is the one *we construct* for ourselves in relationship with others, then the maintenance of narrative agency takes on major importance. (Baldwin 2005. 1025, italics added.)

According to this second passage, *we construct*, or are authors, of the narrative rather than being the narrative itself. Similarly, in the quotation above, Dennett says “In fact *we* have to *posit* selves for ourselves as well” (italics added). But if selves are authors of narratives then what constitutive work is the idea of narrative doing, after all?

Second, if selves are, literally, narratives then how do narratives have meaning? A narrative is made up of a collection of signs (written or spoken). So how can those signs come to carry a meaning? The problem is this. Most plausible accounts of how linguistic meaning is possible presuppose an embodied agent whose beliefs and actions are appealed to to explain meaning. Gricean theories, for example, explain linguistic meaning by appeal to a speaker’s intentions to communicate his or her beliefs by such and such signs (Grice 1969). But a narrative account of self inverts that order of priority and thus must, somehow, explain the meaning of a narrative without appealing to agents. And that seems a difficult venture.

Third, if selves are constituted by narratives then the component elements of the narratives must not presuppose any concept of self. But it is difficult to see how a narrative account of self could avoid including elements which correspond to psychological states. A narrative which avoided all mention of mental phenomena would be useless to explain the notion of a self. But if the narrative presupposes psychological states, then that will surely, illicitly, presuppose a concept of self of whom the psychological states are states (Thornton 2003).

I think that the difficulties Dennett’s account faces are endemic in substantive narrative accounts of the metaphysics of the self. If this is a strand in Schechtman’s narrative account, so much the worse for it.

## 2. *The lives of persons are narrative in form*

Schechtman also invokes a distinct idea: the lives of persons are narrative in form. One philosopher she invokes in support of this idea is Alasdair MacIntyre. This idea, while still contentious, is much less radical than the idea that selves just are narratives. MacIntyre, for example, is led to it through a consideration of the way action explanation iterates.

It is a conceptual commonplace, both for philosophers and for ordinary agents, that one and the same segment of human behavior may be correctly characterized in a number of different ways. To the question ‘What is he doing?’ the answers may with equal truth and appropriateness be ‘Digging’, ‘Gardening’, ‘Taking exercise’, ‘Preparing for winter’ or ‘Pleasing his wife’. Some of these answers will characterize the agent’s intentions, other unintended consequences of his actions, and of these unintended consequences some may be such that the agent is aware of them and others not. What is important to notice immediately is that any answer to the questions of how we are to understand or to explain a given segment of behavior will presuppose some prior answer to the question of how these different correct answers to the question ‘What is he doing?’ are related to each other. (MacIntyre 1981. 206.)

MacIntyre argues that the interconnection between different explanations of actions appealing to narrower or broader contexts has the form of a narrative. Actions are intelligible insofar as they can be fitted into an intelligible narrative and this process iterates. Making sense of one action by citing a broader context of action into which it fits itself presupposes that that broader context makes sense. And hence, MacIntyre concludes, it presupposes a narrative view of the whole of a life, which he connects to the idea of personal identity or selfhood.

What the narrative concept of selfhood requires is thus twofold. On the one hand, I am what I may justifiably be taken by others to be in the course of living out a story that runs from my birth to my death; I am the *subject* of a history that is my own and no one else’s, that has its own peculiar meaning. I am not only accountable, I am one who can always ask others for an account, who can put others to the question. I am part of their story, as they are part of mine. The narrative of any one life is part of an interlocking set of narratives... To be the subject of a narrative that runs from one’s birth to one’s death is, I remarked earlier, to be accountable for the actions and experiences which compose a narratable life. (MacIntyre 1981. 217.)

If Schechtman's narrative constitution account is viewed through the prism of MacIntyre's account, then it can be seen as answering to the richer notion of the nature of the moral self rather than to more austere question of basic identity because, as MacIntyre acknowledges, his is a modest and anti-reductionist account of selfhood or personal identity.

I am not arguing that the concepts of narrative or of intelligibility or of accountability are more fundamental than that of personal identity. The concepts of narrative, intelligibility and accountability presuppose the applicability of the concept of personal identity [...] (MacIntyre 1981. 218.)

The account charts a connection between selves as subjects or agents, their actions and the broader narratives into which these fit. It does not provide an independent way to characterise selves, persons or personal identity.

There is reason, too, to believe that this is the sense of identity that Radden and Fordyce have in mind.

New terms and contrasts have been introduced to separate the more political, and more recent, types of identity discourse. Marya Schechtman speaks of identity as concerning *characterization* (Schechtman 1996) (The identity associated with earlier theorizing in the Lockean tradition she designates *reidentification* identity, in contrast.) [...] The notion of characterization identity captures other presuppositions found in less formal discussions of identity as well, particularly those associated with the politics of recognition. (Radden and Fordyce 2006. 73.)

But does a narrative account of a richer notion of identity, associated with identity politics and identity crises, and on the model of MacIntyre's account, support their radical con-constructivist claims?

There is no reason to think it does. MacIntyre's account is a generalisation from action explanation by reasons concerning which neither constructivism nor co-constructivism seem plausible. The ascription, or avowal, of a reason for action *answers* to the facts about motivation rather than constructing them. First person reports can be sincere or insincere because of that. While first person privilege attaches to reports of motivation, as it does for other aspects of mental life, this is not where the element of truth in constructionism applies to action explanation specifically. That element consists in the fact that *actions* are constituted – created, perhaps – by agents. Decisions are made – constructed, perhaps – by subjects and, if they are practical, executed. But subsequent explanations of actions answer to those facts. Co-construction may thus apply to decisions for action jointly arrived at but that provides no grounds for thinking co-constructivism is true of action *explanations*. And hence there is no reason to think it anything other than a distortion of a narrative conception of the whole

of a life, and hence a self, on MacIntyre's view. Co-constructivism is not only dangerously paternalistic in the case of dementia but it is also implausible as any account of personal identity.

#### IV. A POSITIVE USE OF NARRATIVE IN DEMENTIA?

Despite this sceptical view of narrative as a means of co-constructing identity in the face of dementia, it may, nevertheless have *some* role. Consider the case of the offspring who asks of their elderly and confused father: "Is he *still* my dad? Is he still *him*?" Such a question is not a question about what Schechtman calls basic identity, which might be asked by a long-lost offspring in order to find which unrecognised person in a nursing home, for example, is their father. Rather, if the question is asked of a particular person, whether he is *still* their father, it carries a different sense.

I suggest that it is a question about whether there is still sufficient continuity of character to *count* as the same person in a richer but quite specific sense (see below). This is not simply the characterization question as reconstructed using MacIntyre's narrative concept of selfhood (and certainly not as asking about the basic sense of identity in the style of Dennett's account of narrative centres of gravity). It is asked in a context in which, despite the person concerned being the same person – in the basic sense – who used to be a loving father or fugitive Nazi, or, strangely, both, the ravages of memory loss call into question one sense of (Lockean) forensic identity. In the case where the father is a fugitive Nazi, there *may* be no reason to punish a confused old man who has no understanding of that for which he is being punished. But neither the clear answer: "Yes: he is the same person in the basic sense of identity" nor the answer "No: he should not be held forensically responsible for past actions because he has lost his memory of them" is appropriate for *this* question. And yet it is a question that can seem pressing.

That questions asked with the same words may call for quite different answers is nicely illustrated in a discussion of Wittgenstein's critique of Frazer's *Golden Bough*.

Someone bereaved might exclaim, "Why did she die?" Such a question, uttered in a particular tone and under particular circumstances, would only be taken by the most obtuse as requests for an explanation – as being satisfied by the sort of response appropriate to the question in the context, say, of a coronial enquiry. (Redding 1987. 264.)

In such a case, what looks like a question may not be one at all but an expression of loss clothed in interrogative form. By contrast, in the case I have in mind, I do think that the apparent question "Is he still my dad?" is a genuine question

meaning something like: “Is he still the same person that I remember” but in such a way that the notion of sameness is not captured by either the basic or the forensic senses set out above.

I suggest that the asking of this question in this context stands to those senses in the way that Wittgenstein describes as “secondary sense” (Wittgenstein 1953/2009, part II, sec. xi). The first instance he gives is the attitude most of us have towards words. We feel that a word carries its meaning somehow immediately with it. It can lose this kind of meaning if repeated. Wittgenstein describes this kind of immediate perception of the meaning of a word in isolation as a form of understanding meaning. Since Wittgenstein’s official recommendation is to think of understanding as grasp of a practice, the use of the words ‘understanding’ and ‘meaning’ in the case at hand is not straightforward. It is not a metaphor, however, because nothing can be said to explain why we want to use these words for this kind of experience. But whilst this is not a metaphorical use it is nevertheless a secondary use: one which we find natural given the primary use, but which is discontinuous with, and could not be used to teach, the primary use.

Another example Wittgenstein gives is the use of ‘fat’ in the claim that Wednesday is fat.

Given the two concepts ‘fat’ and ‘lean’, would you be inclined to say that Wednesday was fat and Tuesday lean, or the other way round? (I am strongly inclined towards the former.) Now have ‘fat’ and ‘lean’ some different meaning here from their usual one? – They have a different use. – So ought I really to have used different words? Certainly not. – I want to use *these* words (with their familiar meanings) *here* [...] Asked “What do you really mean here by ‘fat’ and ‘lean’?”, I could only explain the meanings in the usual way. I could not point them out by using Tuesday and Wednesday as examples [...] The secondary meaning is not a “metaphorical” meaning. If I say, “For me the vowel e is yellow”, I do not mean: ‘yellow’ in a metaphorical meaning for I could not express what I want to say in any other way than by means of the concept of yellow. (Wittgenstein 1953/2009, part II §§274–278.)

Secondary sense as introduced is neither a metaphor nor a simile because there is no way to begin to justify the use by articulating the similarity with the literal sense. It is spontaneous extension of the paradigmatic meaning of a word but reliant on that meaning.

Whilst experiencing the meaning of a word or ascribing a width to days of the week may seem to be of limited interest, the Wittgensteinian philosopher Oswald Hanfling argues that the secondary use of words is widespread (Hanfling 1991). In aesthetics, he argues, words such as ‘sad’ applied to music are used in secondary sense. (The music need not make a hearer sad, does not sound like a sad person etc.). In the description of feelings, phrases such as ‘pins and needles’

dles', 'butterflies in the stomach' and 'stabbing pains' are all used in this way. Wittgenstein's own description of "feelings of unreality" in which "everything seems somehow not real" is also secondary.

Wittgenstein himself gives an example of an aesthetic description of a piece of music which seems to be a case of secondary sense:

If I say for example: Here it's as if a conclusion were being drawn, here as if something were confirmed, *this* is like an answer to what went before – then my understanding presupposes a familiarity with inferences, with confirmations, with answers. (Wittgenstein 1967. §175.)

The application to music of language whose primary use concerns the description of arguments lacks a justification through an appeal to objective similarities. Someone who was mystified by this deployment of language need have made no mistake. But that is not to say that for those for whom this extension comes naturally there are no aesthetic justifications for hearing a particular phrase as a conclusion of a previous passage of music or as answering a previous phrase. Such justifications might consist in playing the music with a particular emphasis or stripped of some of its detail. So while there is no justification for the whole extension of a language of argument to music, for those who do find it natural there can be justifications or points of disagreement for specific applications. Nevertheless, even in this case, the justifications can come to an end without agreement and again implying no cognitive shortcoming. One may simply disagree with hearing a theme as the conclusion.

Aesthetic discussions [are] like discussions in a court of law, where you try to "clear up the circumstances" of the action which is being tried, hoping that in the end what you say will "appeal to the judge" [...] if by giving reasons of this sort you make another person "see what you see" but it still "does not appeal to him" that is "an end" of the discussion (Moore 1955. 19).

This combination of (unjustified) spontaneous extension but piecemeal partial justification captures the case at hand. The offspring who asks whether their father is still their father, still the same person, still their 'dad', even though already having answers to the basic question and the forensic question is pressing the notion of sameness in a novel way. 'Same' in this case is used in secondary sense.

In this context, narrative can provide the material for a response which does not simply answer to the facts. Imagine a case in which the elderly person in question cannot recall who his children are but, despite that, he reacts to them with affective warmth. Perhaps hearing a once familiar tune prompts jovial attempts to whistle the refrain. Perhaps he is visibly calmer in the presence of his



offspring, instinctively reaching out a hand. In such a case, a narrative account may weave together these affective responses in such a way to trump cognitive failings. Perhaps the offspring sees in a single gesture, or hears in a single characteristic utterance, sufficient sign of the presence of the remembered father to be reassured and to answer the question in the affirmative.

In such case, the mark of the success of the narrative in conveying an answer to the question – whether happily positive or sadly negative – is not simply answering to neutral facts. Rather, it is the response, to the proffered narrative answer, of the person who asks the question, akin to Wittgenstein's account of aesthetic justification. The success of the narrative answer lies in part at least in the affective response of the person who asks the question. What constitutes the success of the narrative in affirming or denying that the person still is the father is not its objectively successfully marshalling the facts but rather its being accepted as significant by the questioner.

That is, I suggest, the very limited truth in a constructionist approach to a narrative form of identity in dementia, one which carries no risk of insidious paternalism.

#### REFERENCES

- Baldwin, Clive 2005. Narrative, Ethics and People with Severe Mental Illness. *Australian and New Zealand Journal of Psychiatry*. 39. 1022–1029.
- Baldwin, Clive 2008. Narrative, Citizenship and Dementia: The Personal and the Political. *Journal of Aging Studies*. 22. 222–228.
- Dennett, Daniel 1992. The Self as a Center of Narrative Gravity. In Frank S. Kessel – Pamela M. Cole – Dale L. Johnson (ed.) *Self and Consciousness: Multiple Perspectives*. Hillsdale/N, Erlbaum. 103–116.
- Grice, Paul H. 1969. Utterer's Meaning and Intentions. *Philosophical Review*. 78. 147–177.
- Hanfling, Oswald 1991. I Heard a Plaintive Melody. In Allen Phillips Griffiths (ed.) *Wittgenstein Centenary Essays*. Cambridge, Cambridge University Press. 117–134.
- Hume, David 1978. *A Treatise of Human Nature*. Oxford, Oxford University Press.
- Keady, John – Sion Williams 2005. Co-Constructed Inquiry: A New Approach to the Generation of Shared Knowledge in Chronic Illness. Paper presented at RCN International Research Conference. Belfast.
- Locke, John 1975. *An Essay Concerning Human Understanding*. Oxford, Oxford University Press.
- MacIntyre, Alasdair 1981. *After Virtue*. Notre Dame, University of Notre Dame Press.
- McDowell, John 1998. *Mind, Value and Reality*. Cambridge/MA, Harvard University Press.
- Moore, George E. 1955. Wittgenstein's Lectures in 1930–33. *Mind*. 64. 1–27.
- Radden, Jennifer – Joan M. Fordyce 2005. Into the Darkness: Loosing Identity with Dementia. In Julian Hughes – Stephen Louw – Steven Sabat (eds.) *Dementia: Mind, Meaning and the Person*. Oxford, Oxford University Press. 71–88.
- Redding, Paul 1987. Anthropology and Ritual: Wittgenstein's Reading of Frazer's *The Golden Bough*. *Metaphilosophy*. 18. 264.
- Schechtman, Marya 1996. *The Constitution of Selves*. New York, Cornell University Press.

- Schechtman, Marya 2014. *Staying Alive: Personal Identity, Practical Concerns, and the Unity of a Life*. Oxford, Oxford University Press.
- Thornton, Tim 2003. Psychopathology and Two Varieties of Narrative Account of the Self. *Philosophy Psychiatry and Psychology*. 10. 361–367.
- Thornton, Tim 2005. Discursive Psychology, Social Constructionism and Dementia. In Julian Hughes – Stephen Louw – Steven Sabat (eds.) *Dementia: Mind, Meaning and the Person*. Oxford, Oxford University Press. 123–142.
- Wittgenstein, Ludwig 1967. *Zettel*. Oxford, Blackwell.
- Wittgenstein, Ludwig 1953/2009. *Philosophical Investigations*. Oxford, Blackwell.

GERGELY AMBRUS

## The Identity of Persons: Narrative Constitution or Psychological Continuity?\*

### I. INTRODUCTION

This paper investigates Marya Schechtman's narrativist account of self and personal identity, which she dubbed the "Narrative Self-Constitution View". I lay out the main features of this conception by contrasting it with the views of Derek Parfit, a major contemporary representative of the psychological relationist tradition originating from Locke and Hume, to which Schechtman's theory, and narrativism in general, may be seen as a major challenge. In the discussion I will also refer to some other notions of the self, namely the minimal self conceptions of Dan Zahavi and Galen Strawson which, I take it, are also relevant for the reconstruction and evaluation of the narrativist *vs* psychological relationist debate.

I will proceed as follows. First, I provide a brief summary of Parfit's and Schechtman's account of the nature of persons and of personal identity. Then I discuss some points of Schechtman's criticism of Parfit's view, focusing on memory, and argue that Parfit's notion of q-memory may be saved from Schechtman's objections. As a consequence, Parfit's psychological relationist view of diachronic personal identity, which is elaborated in terms of the notion of q-memory (and q-belief, q-desire and other q-states), need not be discarded necessarily. However, I also argue that Parfit's view, according to which what matters is only the holding of R-relation, is also wanting. I maintain, in contrast to Parfit, that in order relations that matter to hold, identification is necessary. Lastly I discuss possible relations between identification, minimal self and narrative self.

\* This paper is based on research carried out in the frames of the K-120375 NKFI-OTKA research project of the National Research, Development and Innovation Office, Hungary.

## II. SCHECHTMAN AND PARFIT: THE NATURE OF PERSONS AND PERSONAL IDENTITY

### 1. Parfit's reductionist view of persons

According to Parfit's reductionist view of persons:

- (1) A person's existence just consists in the existence of a brain and a body, and a series of interrelated physical and mental events.
- (2) A person is an entity that is *distinct* from a brain and a body and such series of events, that *has* a body and brain, and *has* thoughts, desires etc. But it is *not a separately existing* entity. This view about the relation of persons and their mental states and their relations may be characterized as ontologically reductionist but conceptually non-reductionist.<sup>1</sup>
- (3) The facts that determine a particular person's existence can be described in *an impersonal way*, that is, without either presupposing the identity of the person, or explicitly claiming that the experiences in the person's life are had by that person, or even without explicitly claiming that that person exists. (This amounts to the claim that particular mental states can be identified without reference to their subject, the person who "has" them. Cf. Parfit 1984. 189 ff.)

The identity of a person through time, according to the psychological relationalist view, is constituted by *psychological continuity and/or connectedness*,<sup>2</sup> what Parfit calls, following Russell, the 'R-relation'. *Psychological connectedness* is the holding of direct psychological connections. *Psychological continuity* between two persons existing at different times is the holding of overlapping changes of strong (more than 50%) psychological connectedness between them. A person  $P_2$  at  $t_2$  is identical with person  $P_1$  at  $t_1$ , if they are *psychologically continuous and/or connected*, and *there is no other* person  $P_2^*$  at  $t_2$ , with whom  $P_1$  at  $t_1$  is also psychologically continuous and/or connected.

From the above characterization of persons it follows that it is *possible* that the same human being may not be not the same person at different times. Parfit not only acknowledges but welcomes this consequence. He asserts, however, that this does not pose a threat to "what matters", i.e. to what we take personal identity to be important for. What matters, according to him, are the following relations between persons which are essential for personal existence: respon-

<sup>1</sup> Similarly to Hume's account of the existence of nations. According to him, nations are constituted by their citizens, territory, institutions, culture etc., but they are distinct from their constituents in the sense that they have different properties; many predicates of nations cannot be predicated truthfully or meaningfully of their citizens, territory, institutions etc.

<sup>2</sup> Parfit uses slightly different definitions of R-relation in different contexts (see Belzer 1996), but this need not concern us here.

sibility, compensation, survival and self-interested concern. According to the traditional view, a person is responsible for a past action, if he or she is identical with the agent of that action. Similarly, a person is entitled for compensation for past harms of a person if they are identical. Furthermore, a person's survivor is the future person who is identical with him or her; and a person is justified in having a special concern for a future person, if the future person is identical with him or her. All these relations that matter are grounded in the identity of persons. Parfit, in contrast, holds that it is not identity but the holding of R-relation that grounds these relations. A person is responsible for a past person's deed or entitled for compensation for the harms suffered by a past person if they are R-related. And a person's survivor is the future person with whom he or she is R-related; and a person is justified in having a special, self-interested concern for a future person if they are R-related. But R-related persons are not necessarily identical, they may be different persons.

## 2. Schechtman's non-reductionism: the Narrative Self-Constitution View

Central to Schechtman's view is the notion of narrative, which she characterizes, following Bruner, as follows:

A narrative is composed of a unique sequence of events, mental states, happenings, involving human beings as characters and actors. These are its constituents. But these constituents do not, as it were, have a life of meaning of their own. Their meaning is given by their place in the overall configuration of the sequence as a whole – its plot or *fabula*. (Bruner 1990. 43–44.)

More in detail:

To say that a person's life is narrative in character, is at least partly to claim, that *no time-slice is fully intelligible*, or even definable *outside the context of life* in which it appears. [...]

We expect a person's beliefs, desires, values, emotions, actions and experiences to *hang together in a way* that makes what she says and does and feels *psychologically intelligible*. The general gist of this observation can be captured by considering the distinction we recognize between fictional characters who are well drawn and those who are not. Sometimes the collection of actions, thoughts, emotions, and characteristics ascribed to a character make sense – we can understand her reactions, motivations and decisions – they pull together a robust picture. Other times, however, we are at a loss to put together the information we are given about a character.

[...]

A parallel distinction can be drawn in the case of biographical and autobiographical narratives. There are stories of lives, and the subjects of these stories can be well-defined ones just as the protagonists of the fictional narratives can be.

Roughly, then, the narrative self-constitution view requires that a person have a self-conception that coheres to produce a well-defined character. (Schechtman 1996. 96–97.)

Thus Schechtman's conception of personhood essentially relates to the concept of narrative. According to her view:

- i. A person emerges when and by a conscious being begins to conceive himself as a persistent entity constant through time; and this attitude is grounded by interpreting his or her experiences, acts and events of life into a *narrative*. The – emergent – person, who is a “product” of organizing memories, experiences and other mental states into a narrative, is logically prior to his or her experiences, hence cannot be reduced to them.<sup>3</sup>
- ii. A person is a *holistic* complex, his or her mental states are not discrete and cannot be identified atomistically. The contents of experiences, by being organized into a narrative, mutually inform and influence each other. (The “soup” or “stew” theory of the self, as she dubs it).

Schechtman's account of the diachronic identity of persons is also connected to the notion of narrative, and differs fundamentally from Parfit's psychological relationist view. She maintains that

- iii. The self, constituted by experiences narratively organized, is a persistent entity, remaining identical through time. The reason is that when a conscious being conceives his experiences and life events in a coherent narrative, he or she *eo ipso* conceives the protagonist of this narrative as a persistent entity.
- iv. Persons are *temporarily extended* entities. A single consciousness over time is constituted by those particular mental states and features instantiated at different times which mutually influence and inform each other. They are part of the same person because they belong to the same narrative.

Schechtman's account of what matters, i.e. responsibility and the other three features, is also tied to narrativity. She holds, in line with the traditional view, that a person is responsible for a past deed, if he or she is identical with its agent,

<sup>3</sup> This may sound circular, since, according to i., the person *emerges from* organizing *experiences* (memories of past experiences and acts), how could then the person be logically prior to experiences? This objection may be avoided by the assumption that the phenomenological nature of the experiences change by being arranged into a narrative and thereby experiencing them *as the* (same persisting) *person's* experiences.

if they are the same person. However, according to her narrativist approach, being the same person is accounted for as the agent's being the protagonist of the narrator's life narrative. So, according to Schechtman, Parfit's "re-identification question", i.e. the formulation of the question of responsibility for past actions as a question about the re-identification of persons, is misguided (and similarly concerning the other features that matter). A person who emerges by conceiving himself as the protagonist of his life narrative is *eo ipso* persistent, therefore it does not make sense to ask whether some criterion is met which would *ground* the re-identification of the person with the agent of the a past action. Instead, what is relevant concerning responsibility (and the other features) is the "characterization question" addressing which experiences, beliefs, emotions, psychological character traits constitute a particular person; and the issue whether an act may or may not be credited to the person ought to be based on this. The re-identification question is also relevant, according to Schechtman, but it concerns the identity of the *body*. Re-identifying persons via their bodies constrains but does not determine the kind of psychological configurations that constitute a single psychological subject.

### III. SCHECHTMAN'S CRITICISM OF PARFIT

#### 1. *Schechtman's objections against Parfit's reductionist view*

According to Schechtman, the criterion of diachronic re-identification of persons proposed by Parfit relies on the following assumptions:

- i. Persons (i.e. the mental states which constitute them) are impersonally identifiable.
- ii. Particular mental states are atomistically/discretely identifiable.
- iii. The nature of episodic memory is correctly accounted by the so-called "storehouse-theory" of recollection.<sup>4</sup>

Schechtman rejects all these claims. I will discuss in detail only i. the question whether an impersonal identification of a person is possible. Schechtman's major objection against i. derives from the phenomenology of remembering, but further support is provided by her rejection of ii. and iii. I do not intend to dis-

<sup>4</sup> According to this view, memory is seen as a sort of warehouse in which our ideas and experiences are laid away for later retrieval in their original form. The "storehouse" conception was arguably held by Plato, Augustine, Hobbes, Hume and Locke. Schechtman does not claim that contemporary psychological continuity theories, or Parfit in particular, explicitly embrace the storehouse theory, but she claims that it fits well with their implicit understanding of remembering that grounds their account of diachronic identity of persons. Cf. Schechtman 1996b, 6 ff.

cuss the issue of atomistic identifiability and the storehouse-theory of memory here, but let me note in passing that, arguably, even accepting Schechtman's rejection of ii. and iii. does not necessarily undermine a Parfit-style psychological relationist account of persons and personal identity.

One strategy to show this could be to argue that the holistic nature of the content of mental states is less comprehensive than Schechtman takes it to be. Perhaps a more plausible view is a sort of "molecularism", according to which the changes of particular mental states affect not the whole web but only a smaller set of mental states the contents of which influence each other and liable to change together. Such molecularism does not necessarily contradict the idea that a particular person at a time may be identified by the set of interrelated mental states, and that diachronic identity may be accounted for in terms of psychological continuity and/or connectedness.

As for the storehouse-theory, Schechtman's objections are based on results from empirical research on memory. (See e.g. Barsalou 1988, Ross 1989, Barclay-DeCooke 1988.) According to these, autobiographic memory consists of episodic memories of particular events (experiences or acts) to a much lesser degree than it was earlier supposed; the majority of autobiographical memories are condensed memories of certain experience or activity *types*, with which one was typically engaged in a certain period. Furthermore, many of our episodic memories are constructed, moreover, it is often the case that such constructed memories are literally false: the events the subjects (honestly) seem to remember did not in fact happen. Interestingly, however, these false memories often correctly characterize the nature of the (falsely) remembered real events or situations.<sup>5</sup>

Again, it may be possible that Parfit's view of diachronic identity can be accommodated with the constructive theory of memory. For the essence of his psychological view is that identity is preserved if the process of change in of the overall content of a mind is continuous (i.e. not abrupt) and its pace is relatively slow. But diachronic identity of a person does not only consist in the availability of (veridical) episodic memories but it also involves the persistence of other kinds of mental characteristics, psychological character traits, long-terms goals, moral values and so on, which are not constructed in the manner of episodic memories. Furthermore, even the condensed nature of many autobiographic memories seems to be no hindrance for identifying one's past activity in a certain period of his or her life. The constructedness of episodic memory however, seems to pose a more serious threat. But the extent to which constructedness threatens diachronic identification based on psychological continuity depends on the proportion of distorted memories. Moreover, even having a large number

<sup>5</sup> See e.g. Neisser's discussion of John Dean's testimony at the Watergate hearings about his conversations with president Nixon. Neisser 1990. Cf. Schechtman 1990b. 8 ff.



of false memories, which nonetheless characterize the remembered situation correctly, may not be fatal for a Parfitian memory/psychological continuity theory either.

## 2. *Schechtman's circularity objection*

After these brief remarks about the constructedness of memory and the discrete identifiability of mental states and their connection with the psychological relationist view of personal identity, I turn to Schechtman's objection against the possibility of an impersonal identification of persons.

The argument of Schechtman against Parfit may be seen as a twist on Butler's classical circularity objection to Locke's "memory criterion", according to which:

(It is) self-evident that consciousness of personal identity presupposes, and therefore cannot constitute personal identity, any more than knowledge in any other case, can constitute truth which it presupposes (Butler 1736, in Perry 1975. 100).

Butler claims that it is a conceptual truth that a person can only remember his or her own experiences. Parfit, in reply to this objection, introduces the notion of q-memory (quasi-memory). Accordingly,  $P_2$  at  $t_2$  q-remembers a particular mental event (say, a perceptual experience) with the content  $M$ , if:

- i.  $P_2$  at  $t_2$  seems to remember a mental event, with the content  $M$ .
- ii. There existed a person  $P_1$  who actually had a mental event with the content  $M$  at  $t_1$ .
- iii. The memory-like mental event with content  $M$  of  $P_2$  at  $t_2$  was caused by the mental event with the content  $M$  of  $P_1$  at  $t_1$  by any cause.

Think of the following situation, for example. John and Jane spend their vacation together in Venice. After dinner John goes for a walk while Jane stays at the hotel and falls asleep. John sits down on St. Mark's Square by the water. The weather is stormy: at 11. p.m. John sees a huge lightning vis-a-vis, striking the bell-tower of the church San Giorgio Maggiore.

According to Parfit's suggestion it is conceivable that at some later date Jane q-remembers John's visual experience of that lighting. This means that

- i. John in fact had a visual impression of a lightning that stroke the bell-tower of San Giorgio Maggiore at  $t_1$ .
- ii. Jane seems to remember a particular visual impression of a lightning striking San Giorgio Maggiore similar to John's at  $t_2$ .
- iii. Jane's apparent memory is caused by John's experience in the right way (i.e. by any reliable mechanism, including science-fiction devices).

Schechtman's objection is the following. Many conscious experiences are personal, in the sense that their content is essentially linked to the life of their subject. Schechtman illustrates this by borrowing Edward Casey's characterization of the content of memories. Casey writes:

I recall going to the movie *Small Change* a few weeks ago [...] The lights dimmed, and *Small Change* began directly. (Or was there not a short feature first? – I cannot say for sure.) The film was in French, with English subtitles. I have only a vague recollection of the spoken words; in fact, I cannot remember any single word or phrase, though I certainly remember the characters *as speaking*. The same indefiniteness applies to the subtitles, at which I furtively glanced when unable to follow the French. Of the music in the film I have no memory at all – indeed, not just of *what* it was but *whether* there was any music at all. In contrast with this, I retain a very vivid visual image of the opening scene, in which a stream of school children are viewed rushing home, seemingly in a downhill direction all the way. The other two scenes also stand out in my present recollection: an infant's fall from a window of a high-rise apartment (the twenty-ninth floor?) and the male teacher (whose name along with all others in the film I have forgotten) lecturing passionately to his class about child abuse. Interspersed between these scenes is a medley of less vividly recalled episodes, ranging from fairly distinct (the actions of the child-abusing mother) to quite indistinct (e.g. children's recitations in the classroom). While I am recollecting this uneven and incomplete sequence of filmic incidents, I find myself at the same time remembering my own children's ongoing reactions to the film. I do not remember their behaviour in detail but only as a kind of generalised response consisting of laughing, whispered questions, outright comments, and the like. These reactions are as intrinsic to the memory as the unfolding of the film itself; so too is the mixture of pleasure and exasperation which I felt being located, as it were, *between* children and film. Suddenly my memory of *Small Change* comes to an end; the lights go up, and we leave through a side exit near us [...] (Casey 1987. 25–26.)

Note that the contents of these memories in some way express that it was *the remembering subject*, Casey, *who* was in the movie-theater, as it involves being there with *his* wife and *his* children in the town *he* was living etc. In other words, the phenomenology of his memories are tainted by being *his* memories, by the fact the objects of the experiences he remembers had a special relation to him.

Now, think of the case when someone else, e.g. Jane from our above story, q-remembers Casey's experiences. Then either:

The phenomenal character of these q-memories *does not involve* that some of the objects or persons in the q-remembered experience *are related to the life of Casey*, to the *subject of the experiences* q-remembered. Then the phenomenal character of her q-memory is quite different from Casey's memory: it appears as seen pictures and heard

sounds following each other, for example, of a woman with children, in a movie and in a city, but *not as of Jane's* wife and *her* children, or *as of her* hometown (as she has no wife, children etc.).

or

It *does involve* that some of the objects of the q-remembered experience *are related to the life of Casey*, i.e. *the subject of the experiences* q-remembered. But then the content of the q-memories is in *manifest contradiction* with Jane's other memories and other mental states. Therefore such a q-memory would have a very strange phenomenal character: Jane would recall being in a movie-theater with her wife and children, reading the English subtitles in order to understand the French film, while at the same time she is also aware that she has no wife and children and has a very good command of French... The phenomenal character of this confused state is surely different from that of Casey, who simply remembers his wife and children and the movie.

In sum: either way, the phenomenal character of Jane's q-memory and Casey's memory of the same past event would not be the same. As a consequence, memory cannot be analysed as *q-memory with a normal cause* (bodily continuity of the same human being), therefore Butler's circularity objection is not answered.

Furthermore, the same applies to other sort of quasi-states, such as q-beliefs, q-intentions and the rest: they cannot replace normal mental states. Hence Parfit's contention that it is R-relation that matters (concerning responsibility, compensation, survival and self-interested concern) cannot hold, since R-relation is defined as a relation between persons who are not identical, but whose psychological relations can be interpreted as q-states of the other person's states (for example, a later person can q-remember an earlier person's experiences, or a later person can q-intend an earlier person's intentions, with whom he or she is R-related).

### 3. Reply to Schechtman's circularity objection

Schechtman is right, the phenomenological character of a memory state about one's own earlier experiences is often different from the phenomenological character of a q-memory of another person about the same event, since either the phenomenal character of the q-memory lacks the personal aspect the content of memory has, or it has it, but then it is manifestly inconsistent with the content of other mental states of the q-remember.

But we may change the definition of q-memory somewhat by adding a *further condition*, namely:

iv. A particular q-memory relation can only take place between psychologically continuous persons.<sup>6</sup>

If we accept iv., I maintain, Schechtman's objection does not hold: memory and q-memory experiences of a particular past event will have the same (or similar) phenomenal character. When a *q-memory*

- (1) Has a normal cause (i.e. the persistence of the body),
- or
- (2) Was caused by non-branching replication (one replica comes to existence and the original person dies),
- or
- (3) Was caused by branching replication (i.e. by fission: one or more replicas come to existence and the original person also continues to exist),

then *the q-memory* of the later person(s) *can be consistent and coherent*<sup>7</sup> with the other memories and further mental states of the q-remembering subject.

This is because in case 1) q-remembering is simply remembering – hence if the memory event is consistent and coherent with the other mental states of the subject, then so is the q-memory event. In cases 2) and 3), non-branching and branching replication, if *the new environment* (the space, time and social environment into which the replica is “born” and where he or she continues the life of the original person) is relatively “close” to *the original*, then the replica's mental contents will be largely *consistent and coherent* with the content of his or her q-memory. *If the new environment is very “distant”*, spatially or temporarily or socially, – for example, the replica emerges on an alien planet, or at a much later date, or in a very different social environment, (as it happens, for example, in Mark Twain's *The Prince and the Pauper*), then the q-memory will be *strongly inconsistent and incoherent* with the content of other mental states of the replica. To put it another way: *if the trajectory* leading from the subject of the remembered experience to the current replica *is not (too) gappy*, then the *consistence and coherence* of the mind of the replica *prevails*.

It may be objected, however, that a *q-memory may be inconsistent and incoherent* with other memories and further mental states of the q-remembering subject, *if he or she is a fusion* of two (or more) past persons, q-remembering one of his or her predecessor's experience. This is true, but *irrelevant*. The reason is that iv., the requirement of psychological continuity, is not met in the case of fusion. If

<sup>6</sup> I understand Parfit's original formulation of psychological continuity as a relation that allows for temporary gaps in the existence of the psychological continuous persons. In my view, this is in line with Parfit's intended understanding of replication and teletransportation.

<sup>7</sup> At least to the degree in which the mental states of a normal person are consistent and coherent.

the content of John's and Jane's mind are fused, then the resulting mind would be inconsistent and incoherent to a great degree. But the relation between the minds of Jane and the fused person "John-and-Jane" would not be psychologically continuous, in Parfit's sense, because there would be no overlapping chains of strong psychological connectedness between Jane's mind just before the fusion and John-and-Jane's mind right after the fusion. (Except in cases when the contents of mind of the predecessor persons to be fused, John and Jane, were very similar just before the fusion. But then, the mental contents of the fused-person would *not* be inconsistent and incoherent.)

Admittedly, condition iv., by which we can answer Schechtman's objection, is not included in Parfit's original definition of q-memory. For example, Jane at  $t_2$ , when q-remembering the experience of the lightning John saw, and John at  $t_1$ , when seeing the lightning, are not supposed to be psychologically continuous. But, I believe, condition iv. seems to conform to the general spirit of psychological continuity theories. In my view, psychological continuity theories were motivated by providing an account of personal identity, which, on the one hand, emphasizes the importance of the *psychological relations* between persons, but does not appeal to the concept of a persistent *mental substance*, while on the other, also emphasizes that – although *bodily continuity* is an important aspect of personal identity – *numerical identity of the body* is not a necessary requirement.

Furthermore, the motivation of the early psychological continuity theorists was supposedly *not* to propose an account, according to which numerically *different human beings, with different bodies and different life histories* (which evolved in different spatio-temporal and social environments), i.e. persons who are not psychologically continuous, *could be smoothly psychologically connected* so that the resulting minds would be "normal", that is, similar to minds having psychologically connections only with persons psychologically continuous with them. In my view, this latter understanding of fusion (which seems to be supported by some writings of Parfit, see e.g. Parfit 1971) goes beyond the original (and, to me, legitimate) ambitions of psychological continuity theories of personal identity.

#### *4. A variant of the circularity objection based on immunity to error through misidentification (IEM)*

According to some philosophers, for example, Wittgenstein and Shoemaker, memory-states are immune to error through misidentification (IEM). (Cf. Wittgenstein 1958. 66–67; and Shoemaker 1968.) This means that a subject cannot err about who is the subject of a particular memory, i.e. whether it was he or she or some other person the experience of whom he or she remembers or seems to remember.

This is not to say, of course, that we cannot have false memories. A memory can be false if its content is not veridical. This may happen in several ways. For example, an event of recollection may be of an experience that really took place in the past, and the content of the experience was veridical at the time of having it, but later recollection of the original experience is distorted. Another possibility is that the content of an experience was non-veridical already in the original situation, and this false content is recalled correctly. Or there was no such experience at all, which the subject seems to remember, but it was made up entirely by him. (Either stimulated by external causes, as in the case of certain law processes, where criminal suspects make up false memories under the influence of the investigators; or by internal causes, when a subject expects that the event he or she seems to remember must have happened, based on his or her other beliefs and desires.) So memories can be false.

But they cannot be false because of misidentifying their subject. Proponents of the view that memory is immune to error through misidentification argue, that we identify ourselves as subjects of mental states in a special way (and perhaps also as subjects of bodily states that when we experience as states of *our body*). We identify ourselves as subjects in a direct and infallible first-personal way, differently from the way we identify objects. Such direct, first-personal identification of one's self is at work also in the case of remembering. Such identification of the self does not assume any ascriptive knowledge about the self; identification through description applies only when we identify ourselves as objects.

Thus the objection against the psychological relationist view based on the immunity-claim is as follows. Since it is impossible to err about the subject of the remembered experience, therefore it simply makes no sense to identify a past person with a present one based on the holding of R-relation between them. For this would be a case of ascriptive identification, i.e. a mode of identification when there is a prior description of the nature of one's self and identification would consist in checking whether the past person satisfies this description.

##### *5. Reply to the IEM-based circularity objection*

One may reply to this objection as follows. If fusion were a real possibility then mistaking the subject of a remembered experience seems conceivable. Clairvoyance may be another example. Clairvoyance may be seen as a kind of q-remembering, because telepathic "seeing" of some past experience of another person satisfies the definition of q-memory. And it seems conceivable that a clairvoyant person misidentifies the subject of a mental state he or she directly experiences. Such cases could occur when the "seer" would mistake a recollection of his

or her previous mental state for a veridical clairvoyant state about some other person's experience. Since such cases are conceivable, it is not necessary that a person could not be mistaken about who is the subject of some experience he or she remembers or q-remembers. Therefore, I maintain, this sort of circularity objection is not successful. It is not true that psychological continuity cannot establish personal identity because a person can only be psychologically continuous with oneself, since a person can have memory-like states only about his or her past experiences, as one cannot err about the subject of experiences he or she seems to remember.

#### IV. THE INEVITABILITY OF IDENTIFICATION

As already discussed, according to the traditional view personal identity is required for what matters. According to Parfit's alternative conception, what is important for what matters is not identity, but the holding of R-relation. If two persons are R-related, then the relations of responsibility, eligibility for compensation, survival and self-interested concern hold between them. If a person is not R-related to another (even if they are bodily continuous), then ascribing responsibility to him or her for the acts of the other person is not justified (and similar considerations apply to the other three features).

My view differs from both. I hold, in opposition to the traditional view and in line with Parfit, that identity is not a matter of fact. But I also hold, in opposition to Parfit, that the holding of the R-relation is not sufficient for what matters. In contrast with both, I hold that identity is necessary for what matters, but identity is established by identification.

Thus identification is necessary for what matters. Merely the holding of R-relation is not sufficient, for consider: would *simply* the fact and my knowledge of it that the mental states of some other past person were very similar to mine, i.e. we are "psychological twins", as it were, would bring about my taking responsibility for that other person's deeds? And similarly: would the mere fact and my knowledge of it that some other person is my psychological twin induce a special concern in me for his or her well-being? Or, were I to die, would I then believe that my psychological twin is my survivor? I do not think so. In order to take responsibility for a past person's deeds, I have to identify myself with that person. But merely being strongly psychologically connected or even psychologically continuous with a person seems not sufficient for identifying myself with him or her. And the same applies to survival and self-interested concern: strong psychological connectedness or even psychological continuity seems not sufficient for considering a future person to be my survivor or to be worthy of special, self-interested concern. I have to identify myself with that future person to view him as my survivor who deserves my special concern.

My view may also be formulated as claiming that identity does matter, but identity is created by identification, it is not a pre-existing fact which is recognized by the act of identification.

## V. WHAT GROUNDS IDENTIFICATION?

If identification is necessary for what matters, the question then emerges: what is the origin of identification? Here are some possible options.

*R-relation.* Identification may be founded merely upon being R-related: simply, if a person is R-related to another person this fact may itself bring about it that he or she identifies with the other. However, this does not always seem to be the case. Even in the normal course of events it sometimes happens that a person does not identify him- or herself with the agent of a past deed with whom he is R-related. There are also pathological cases in which subjects do not identify (or even explicitly reject identifying) themselves with the subject of “their” past actions. And there are also conceivable cases of replication when it seems plausible to assume that the replica may not identify him or herself with the “original” person, or that a person does not identify with a future replica. Such cases may be when the environments of the replica and the original person differ significantly (i.e. they are very far from each other, temporarily or spatially or socially).

*Minimal self.* Another option may be to assume that there exist minimal selves in Zahavi’s sense (cf. Zahavi 2006, 2010), and identification is based on the identity of our minimal self. According to Zahavi, all episodic memory-states involve an “elusive sense of presence”, i.e. the content of memory experiences involve the feeling that the subject of the remembered experience was *me*, the same person as the remembering subject. This supposed phenomenological feature of memory experiences Zahavi terms as (having) a minimal self. A minimal self is “transcendent in immanence” as Husserl put it (Husserl 1976. 123–124): it does not exist beyond the content of conscious experience, but it is permanently present within all conscious experiences (cf. Zahavi 2011. 327–328).

If we accept this view, a further question may also arise concerning the origin of minimal self. One option is that certain *relations*, namely the holding of R-relation between (the mental states of) different persons cause the emergence of the minimal self, our sense of diachronic unity. A person identifies with the subject of a remembered experience, because he has a sense of being the same subject.

There are several arguments against such reductionist views, which aim to explain the feeling of being the same subject over time in terms of certain *relations* between certain mental states, instantiated at different times. Hume, for example, suggested in the *Treatise* that the illusion of a persisting self is creat-



ed by psychological mechanisms based on the similarity of content of, and the causal relations between mental states. Our mind tends to judge that mental states (perceptions) that have similar content are about the same object. (Hume 1739/2007. *Of Personal Identity*. Book I Part IV, Section VI.) Therefore, if the mental states of a person remembered are similar to a sufficient degree to the mental states of the remembering person this automatically induces the belief in the rememberer of being identical with the remembered person. However, as it is well-known, Hume revoked this proposal in the *Appendix*, and admitted not being able to conceive how *relations* between completely *distinct* (ontologically independent) mental states could be the source of the belief in the existence of having a persistent self (Hume 1739/2007. *Appendix*).

Husserl, and following him Zahavi, formulated another objection against such relationist views (cf. Zahavi 2005. 49–72). According to objection, the minimal self cannot be constructed out of the particular contents of consciousness and relations among them, because the synchronic unity of consciousness presupposes time-consciousness, which, in turn, presupposes the diachronic identity of the self. The reason is that time-consciousness is rooted in the structure of conscious experiences. Conscious experiences are not momentary; they are more like blocks or fields: beyond the actual experience the experience just passed is still retained in consciousness (retention), while we also have expectations about the upcoming next experiences (protention). But such triadic structure to be possible, the subject of experience must be identical, i.e. unchanging in which the stream of experience sets forth (Husserl 1952. 98; 1974. 363). Hence time-consciousness, awareness of the passage of time, requires the diachronic identity of the subject. Therefore, the idea that the holding of R-relation may create the sense of diachronic identity is ruled out.

If these arguments are accepted, then the minimal self is not reducible to R-relation between experiences instantiated at different times. Still, we may accept the reality of irreducible minimal selves which could ground identification just as well.

Zahavi also characterizes the concept of minimal self in comparison with Albahari's notion of perspectival self (Zahavi 2010). Albahari proposes a distinction between perspectival self and higher self, and correlatively between two notion of consciousness, witness-consciousness and ownership-consciousness (Albahari 2009). Albahari's perspectival self is similar to Zahavi's minimal self in many, though not all, respects. Having a perspectival self is tantamount to being witness-conscious of all experiences. Where being witness-conscious means being simply presented with the objects of consciousness. The perspectival self, however, is different from the higher self. The higher self emerges only when one appropriates some of the mental states presented to witness-consciousness; in other words, when one identifies with certain beliefs and desires contemplated by the witness-consciousness. The awareness of our higher self involves a felt

difference between awareness of the self and not-self, while witness-consciousness does not involve such a differentiation.

Zahavi's view is similar, but he holds that already the minimal self is personal (individualistic). This difference between Zahavi's and Albahari's accounts has implications for my view which holds that identification is necessary for what matters. Having only a perspectival self involves no identification, thus it is not sufficient for grounding what matters. Having a higher self involves identification by definition, thus, in Albahari's framework, having a higher-self is required for what matters. In contrast, Zahavi's minimal self is personal, and, since awareness of one's minimal self is pre-reflective, having a minimal sense, i.e. a sense of diachronic identity, does not require an act of identification. But having a minimal self may underlie identification and may be sufficient for what matters.

*Episodic persons.* There are also views that deny the reality of persistent minimal selves. According to Galen Strawson, for example, the sense of diachronic identity is not a universal feature of all conscious experiences (of normal adult humans); there exist also so-called "episodic characters", he himself being one. Such persons do remember their past experiences, but (at least most of the times) do not feel a personal presence, i.e. that they do not feel to be identical with the subject of the remembered experience.

I will not want to discuss this view, for it differs fundamentally both from Schechtman's narrativist and Parfit's psychological continuity view, which I contrast in this paper. Schechtman and Parfit differ in many fundamental points, but they share the assumption that it is a general feature of the self-conception of (normal adult) humans that they conceive themselves as identical persons through their lives. They have different views on what they take to be the origin of this belief (i.e. having a life-narrative or psychological continuity), and also concerning whether it is morally good that we have this belief. Parfit proposes that it would be morally preferable if we got rid of our conviction of having a persistent self through our life. Schechtman is probably neutral on this, as she accepts only the psychological, not the moral narrativity thesis. According to the former, human life as a matter of fact has a narrative form, while the latter, advocated for example by Alasdair MacIntyre,<sup>8</sup> also asserts that a life in search of a (good) life narrative is morally superior to a life that has no such aim. But, again, both Parfit and Schechtman assumes that it is a fact that we conceive ourselves as being the same person through our life. Strawson, however, denies this common ground, and, accordingly, explains responsibility and the other three features in an alternative framework.

<sup>8</sup> See e.g. MacIntyre 1984. chapter 15. 202–215.

## VI. CONCLUSIONS

By the above considerations I hope to have shown that:

- (1) Schechtman's attack on the conception of re-identifying persons based on the holding of R-relation between them may not be successful. In particular, Parfit's psychological relationist view *can* be defended from Schechtman's objection that the notion of q-memory is inconsistent.
- (2) Nonetheless, contrary to Parfit, the relations that matter (responsibility, eligibility for compensation, survival and self-interested concern) cannot be grounded exclusively by the holding of R-relation.
- (3) The relations that matter require identification.
- (4) Identification cannot be not grounded merely in the holding of R-relation.
- (5) Identification may be grounded in having a minimal self.
- (6) Assuming the reality of minimal selves, we may conceive the relation between minimal self and narrative self in two different ways. It may be the case that a life-narrative is constituted only by such experiences, with the subject of which the narrator identifies him- or herself (i.e. only by experiences about the narrator feels a personal presence). Alternatively, it may be the case that the feeling of personal presence emerges by "inserting" the experience into the narrative.<sup>9</sup> If the former is the case, then having a minimal self is a precondition of having a narrative self. If the latter, then the narrative self is prior to the minimal self, as it is the setting up of a narrative which induces the sense of diachronic identity. And, as identification is grounded by the minimal self, it also follows that narrativity is required for what matters.

## REFERENCES

- Albahari, Miri 2006. *Analytical Buddhism: The Two-Tiered Illusion of Self*. New York, Palgrave Macmillan.
- Albahari, Miri 2009. Witness-Consciousness: Its Definition, Appearance and Reality. *Journal of Consciousness Studies*. 16/1. 62–84.
- Barclay, Craig R. – Peggy A. DeCooke 1988. Ordinary Everyday Memories: Some of the Things of Which Selves Are Made. In Ulrich Neisser – Eugene Winograd (eds.) *Remembering Reconsidered: Ecological and Traditional Approaches to the Study of Memory*. Cambridge, Cambridge University Press. 91–126.
- Barsalou, Lawrence W. 1988. The Content and Organisation of Autobiographical Memories. In Ulrich Neisser – Eugene Winograd (eds.) *Remembering Reconsidered: Ecological and Traditional Approaches to the Study of Memory*. Cambridge, Cambridge University Press. 193–244.
- Belzer, Marvin 1996. Notes on Relation R. *Analysis*. 56. 1. 56–62.

<sup>9</sup> To argue in more detail for either of these positions requires further investigations that I cannot undertake here.

- Bruner, Jerome 1990. *Acts of Meaning*. Cambridge/MA, Harvard University Press.
- Butler Joseph 1736. Of Personal Identity. The Analogy of Religion. Appendix I. In John Perry (ed.) *Personal Identity*. Berkeley, University of California Press, 1975.
- Casey, Edward C. 1987. *Remembering. A Phenomenological Study*. Bloomington and Indianapolis, Indiana University Press.
- Hume, David 1739/2007. *A Treatise on Human Nature*. Oxford, Oxford University Press.
- Husserl, Edmund 1952. *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie, II. Phänomenologische Untersuchungen zur Konstitution* (Husserliana 4). The Hague, Martinus Nijhoff.
- Husserl, Edmund 1973. *Zur Phänomenologie der Intersubjektivität: Texte aus dem Nachlass, III. 1929–1935* (Husserliana. 15). The Hague, Martinus Nijhoff.
- Husserl, Edmund 1976. *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie, I* (Husserliana. 3/1–2). The Hague, Martinus Nijhoff.
- Parfit, Derek 1971. Personal Identity. *Philosophical Review*. 80/1. 3–27.
- Parfit, Derek 1984. *Reasons and Persons*. Oxford, Oxford University Press.
- Parfit, Derek 2003. Why Identity is Not What Matters. In Raymond Martin – John Barresi (eds.) *Personal Identity*. Blackwell. 115–143.
- Ross, Michael 1989. Relation of Implicit Theories to the Construction of Personal Histories. *Psychological Review*. 96/2. 341–357.
- Schechtman, Marya 1990. Personhood and Personal Identity. *The Journal of Philosophy*. 87/2. 71–92.
- Schechtman, Marya 1994. The Truth about Memory. *Philosophical Psychology*. 7/1. 3–18.
- Schechtman, Marya 1994. The Same and the Same: Two Views of Psychological Continuity. *American Philosophical Quarterly*. 31/3. 199–212.
- Schechtman, Marya 1996. *The Constitution of Selves*. Ithaca, Cornell University Press.
- Schechtman, Marya 2011. Memory and Identity. *Philosophical Studies*. 153/1. 65–79.
- Shoemaker, Sydney 1968. Self-Reference and Self-Awareness. *Journal of Philosophy*. 65/19. 555–567.
- Strawson, Galen 2004. Against Narrativity. *Ratio*. 17/4. 428–452.
- Strawson, Galen 2011. The Minimal Subject. In Shaun Gallagher (ed.) *The Oxford Handbook of The Self*. Oxford, Oxford University Press. 253–278.
- Wittgenstein, Ludwig 1958. *The Blue and Brown Books*. Oxford, Basil Blackwell.
- Zahavi, Dan 2005. *Subjectivity and Selfhood: Investigating the First-Person Perspective*. Cambridge/MA, MIT Press.
- Zahavi, Dan 2006. Thinking about (Self-) Consciousness: Phenomenological Perspectives. In Uriah Kriegel – Kenneth W. Williford (eds.) *Self-Representational Approaches to Consciousness*. Cambridge/MA, MIT Press. 273–295.
- Zahavi, Dan 2010. The Experiential Self: Objections and Clarifications. In Mark Siderits – Evan Thompson – Dan Zahavi (eds.) *Self, No Self?: Perspectives from Analytical, Phenomenological, and Indian Traditions*. Oxford, Oxford University Press. 56–78.
- Zahavi, Dan 2011. Unity of Consciousness and the Problem of the Self. In Shaun Gallagher (ed.) *The Oxford Handbook of the Self*. Oxford, Oxford University Press. 316–337.
- Zahavi, Dan 2017. Thin, Thinner, Thinnest: Defining the Minimal Self. In Christoph Durt – Thomas Fuchs – Christian Tewes (eds.) *Embodiment, Enaction, and Culture: Investigating the Constitution of the Shared World*. Cambridge/MA, MIT Press. 193–199.

JUDIT SZALAI

## The ‘Reasons of Love’ Debate in Analytic Philosophy: Reasons, Narratives, and Biology\*

### I. INTRODUCTION

In English-language philosophy of the past decades, discussions of the “reasons of love” have revolved around a perceived theoretical dilemma. Do we love persons for (some of) their properties – and have reason to love any another person who also possesses those qualities, or, even better, “upgrade” to someone who possesses them to a greater degree? Alternatively, do we love people without reasons, purely “for themselves” – whatever that means, those “selves” being hard to characterize without recourse to properties. To complement the question of reasons/no reasons, more recently, an historical dimension has been added to the debate: other than relating to a person or her properties in a synchronic way, the joint history of the two partners also plays a role in their mutual attitudes.

In this paper, I will take a step back to look at some of the assumptions behind the debate itself and present a more complex picture, based on distinctions between different forms of romantic love.<sup>1</sup> Concerning the discourse itself, I’ll propose, first, that the “reasons-based”/“no-reason” views do not constitute a genuine theoretical dilemma: we do not love persons for either abstract properties that several individuals can share, or for some elusive “ipseity”. Second, the debate is saddled with a kind of descriptive/normative ambiguity, between why persons love or why they *should* love. Third, many discussions equivocate on the different meanings of ‘love’. Making the relevant distinctions – which are to a significant extent based in biology – advances matters a great deal, as ‘reasons’ apply to different forms of (romantic or quasi-romantic) love in different ways.

\* The paper was presented as a Benda lecture at Károli Gáspár University, Budapest, in May 2020. I would like to thank the participants, especially Gábor Boros, for their remarks, as well as Ronald de Sousa and Glenn Most for their generous comments. The paper is an output of the project *Self-Interpretation, Emotions, Narrativity* (K120375, National Research, Development and Innovation Office, Hungary).

<sup>1</sup> In using Helen Fisher’s work in addressing the reasons of love, I am following Ronald de Sousa (2016), who, to my knowledge, has been the first to explicitly suggest that different forms of (romantic) love relate to reasons in different ways.

The most interesting (and perhaps familiar) cases are those in which different types of factors (personal properties of the beloved as reason-giving, joint history, and bio-psychology) are at odds with each other (e.g., history vs. properties as reasons, history vs. biology, properties as reasons and history vs. biology). Towards the end of the paper, some such possibilities will be explored.

## II. LOVE AS AN EMOTION

The kind of love in the focus of the debate addressed here is *romantic* love. Many of the claims made about this apply to other forms of love as well – primarily, to love of friends. Parts of the similarities are due to the relatively small number of targets, in both cases “chosen” rather than given. In biological or family relations, the reasons of love problem do not arise in the way it does in romance or friendship. Being a child or a parent is normally considered sufficient reason for love, and – barring special circumstances, like severe post-partum depression or parental abuse – failing to love a child or a parent occasions moral censure.

Before attempting to settle the issue of whether love has reasons, we need to say something – however vague and approximative – about reasons. According to Thomas Scanlon’s well-known understanding, reasons are considerations that “count in favour of” an action or attitude (Scanlon 1998). If your wallet is being stolen in front of our eyes, you have a reason to try to get it back (action) and a simultaneous reason to become angry (attitude). Reasons, as opposed to mere causes, not only explain actions or attitudes, but also justify them. Having had too much coffee and too little sleep may explain an outburst of anger, but it certainly doesn’t justify it. Emotions and manifestations of emotions *can*, at least sometimes, be justified, though, as the practice of asking people to account for their affects and attempts to provide such justifications suggest (Smuts 2014, 507). Is this the case with love as well? Do we expect and provide justification for loving someone?

The answer in part depends on whether we consider love an emotion. While it may seem obvious that we should, there are legitimate contrary considerations (de Sousa 2015). Our position will naturally also depend on how we understand emotions. For Paul Ekman, one of the most influential affect psychologists, paradigmatic emotions such as fear, disgust, and anger, are universal, short-term responses associated with characteristic facial and other physical manifestations, feelings, and behavior. If we think of emotions as short-lived biological responses, love, which typically lasts longer than a few minutes, and involves no distinctive phenomenology and no typical facial expressions for much of its duration, hardly fits the bill.

Another fact that speaks against regarding love as a (paradigmatic) emotion is the difficulty of finding a so-called ‘formal object’ for it. The formal object of an

emotion is a property of the target (*viz.*, whatever the attitude is directed at) that is present in all cases in which the affective response is apt. The formal object of fear is *the dangerous*, so in every case in which fear is an appropriate reaction, it is a reaction to the dangerousness of its object. In principle, the number of formal objects corresponds to that of emotion types, that is, every kind has its own formal object. Now what sort of property would love respond to in its target? The candidate most often discussed is 'lovability' – an obviously vacuous and weak one, for whether we love someone or not has little to do with their being 'lovable.'

Both considerations against love being an emotion are grounded in assumptions that are far from uncontentious. The conception of emotions as biologically based affect programs can be considered too narrow, ignoring the more complicated and often more "cerebral" ones like relief, *Schadenfreude*, or perhaps love itself. And lack of a formal object is only relevant to those who subscribe to the idea that emotions grasp some sort of evaluative properties. Several emotion theorists would be quite comfortable with the idea that many affective states cannot be associated with an identifiable kind of characteristic in their targets (e.g. those philosophers who believe in a plethora of "nameless emotions," like Peter Kivy (2014) or Sue Campbell (1998)). Not having to share the underlying assumptions, let's adhere to the traditional understanding of love, according to which it is an emotion, to assess the merits and demerits of the "reason-based" and "no-reason" views.

### III. THE "REASON-BASED" AND "NO-REASON" VIEWS OF LOVE

At first blush, love has a lot to do with the characteristics of the loved one. From the potential lover's perspective, the target has to be attractive, a determinable property constituted by different determinates for different persons (physical attractiveness (further determinable), intelligence, kindness, etc.). Since the relevant properties may vastly differ (intelligence is attractive for sapiosexuals but might be repulsive for some others), the reasons for attitudes provided by those properties of the target are not universal. Also, since no checklist of properties can secure love, those reasons are only *pro tanto*. Still, it is quite intuitively a requirement to be able to name certain qualities of the beloved that drew the lover's attention to them and which are perhaps considered necessary by the lover to maintain the attitude of love.

According to the alternative, "no reason" view, we don't love persons for their properties but "for themselves." The best-known advocate of this position is Harry Frankfurt. In Frankfurt's view love does not respond to value properties, but itself bestows value on the beloved (Frankfurt 2004). While this position, it tends to be pointed out, resolves some undeniable difficulties of the "rea-



son-based” view, it is unfortunately hard to even make sense of it. What is the self abstracted from its properties and how do we have access to it? What is that unchangeable core that is not subject to time and perspective? Rather than trying to grasp this “pure ipseity,” let’s see the arguments against each position. Interestingly, the same arguments, based on commonly held attributes of love, tend to be cited against both. These attributes are the following: (1) Exclusivity: the lover is supposed to focus his attentions on a single person. Failing to do so makes the character of the emotion doubtful. (2) Irreplaceability: switching between or “upgrading” targets again questions the kind of emotion we are dealing with. (3) Non-arbitrary grounds: the beloved person would “object to” being loved “arbitrarily”. (4) Permanence: love is not supposed to last only a short while; it allegedly “alters not when alteration finds.” Can the rival views accommodate these characteristics?<sup>2</sup>

As it has already been noted, the “reason-based” view seems to be incompatible with (1) and (2): If we love persons for particular properties constituting reasons, those very same reasons may induce us to love others too, or to switch to others who have the same qualities to a greater degree. (There is always someone wittier, more muscular, or with eyes of a deeper blue.) As to (3), non-arbitrary grounds: Why should it be exactly those properties selected by the lover for which we are loved? Do those properties have sufficient relation to the characteristics we cherish most or find most essential to ourselves? (Are we comfortable with being loved for our tiny ears or that peculiar way of pronouncing ‘r’-s?) (4) also seems jeopardized by the “reason-based” view: if those particular reasons for loving someone cease to hold, why would the attitude continue to be present? (Persons may lose not only their wit and muscle, but significant parts of their personalities as well, turning disillusioned and sour, demented, etc.)

Concerning the “no reason” view, if we don’t love a person for any particular reasons, the attitude does not seem to admit of any account; loving someone “for themselves” is merely a “just so” explanation. As far as reasons go, loving and not loving that person are on a par: contrary to (3), loving that person is, in this sense, arbitrary. If so, as against (1) and (2), the lover may as well switch to another target. Similarly, unaccountable love may come and go, threatening (4), the permanence of the attitude.

Thus, both accounts seem to fail to accommodate some basic characteristics attributed to love. This might be read as a criticism of these attributions. (Indeed, when we come to the biologically-based psychological differences between different forms of love, holding love as a generic category up to the strict standards of exclusivity, irreplaceability and permanence will prove to be somewhat illusory.) It may also be the case that neither account is adequate. I will

<sup>2</sup> Especially the exclusivity and permanence criteria may be called to doubt; I am not going to address such doubts here.



argue for this conclusion and present a different 'reason-based' view, one that better preserves the characteristics assigned to love, with the limitations that will be noted when coming to its different forms.

In the "reasons of love" debate, properties are generally interpreted as abstract universals, instantiable in a number of objects. This interpretation gave rise to the problems of exclusivity and irreplaceability: if you love a person for their funniness and blue eyes, the instantiation of these properties in other persons will also constitute reasons to love those others. Properties can also be understood in a different way, however: as individual or particular, instantiated in a particular object or person (in ontological parlance, as "tropes").<sup>3</sup> On this understanding, the reason for loving someone is not the fact of their instantiating the abstract properties 'being blue-eyed' and 'being funny,' but the particular blue-eyedness and funniness that exclusively belongs to them. While many individuals may share the same abstract property, they cannot have it in the particular way the beloved person does: I cannot have your blue-eyedness and funniness, and you cannot have mine.

What about non-arbitrary grounds and permanence? Persons would not want to be loved on a whim. For those subscribing to the reason-based view, this means that the lover should not draw a blank when asked about the reasons for her feelings for the beloved. Even if she cannot provide a full explanation, she should be able to say something close enough to the actual properties of the loved one. If she cannot come up with any such consideration, why think that she won't just switch to someone else on another impulse? Reference to properties as particulars, belonging to one person only, meets the requirement of providing justification for loving someone without that justification being capable of being extended to a number of other persons. (Correspondingly, the reasons these properties constitute will be non-universalizable, particular reasons.)<sup>4</sup>

For those supporting the alternative, "no-reason" view, non-arbitrariness amounts to the consideration that the beloved would want to be loved "for themselves." Reference to tropes also helps meet this requirement, underlying, but being distinct from, the "no reason" view. Those properties are instantiated in those particular ways and in those particular combinations in one person only. Thus, appreciating those properties does not imply loving the person for something other than themselves, as the particular set of individual properties is constitutive of the person. This way, persons are loved "for themselves" without being loved for 'no reason.'

<sup>3</sup> Reference to 'tropes' may not be seen as a legitimate solution here by those who see the talk of tropes as a way of trying to evade the crucial distinction between the particular and specific. I owe this observation to Ronald de Sousa.

<sup>4</sup> Concerning non-universalizable reasons, see Dancy 1983.

If persons are so special, their properties being particular rather than universal, why do we often switch to others who do not have those qualities manifested in those particular ways? We should note the bias towards a passive characterization of love in Plato's account of *eros* in the *Symposium*, which has become the cornerstone of replaceability/irreplaceability arguments and its contemporary relations. The "ladder of love," appreciation of the idea of beauty in a person or in many persons, is a variation of the "reasons-based" view. The properties of the beloved impose themselves on us, as it were, and we move on to the contemplation of beauty in its ever higher-level instantiations.

Falling in love is hardly primarily a matter of the properties of any potential object, however. Similarly, switching or upgrading is not primarily a matter of a superior instantiation of those properties in the one switched to. Rather, it is triggered by different needs, convictions, and properties of the lover herself, which make it possible for her to pay special attention to similar and non-similar qualities in persons other than the original beloved. The motivations to cheat and potentially fall in love with a person outside an official relationship would be too numerous to list: "Some seek attention. Some want autonomy. Some want to feel special, desired, more masculine or feminine, more attractive or better understood. Some want more communication, more intimacy, or just more sex. Some want to solve a sex problem. Others crave drama, excitement, or danger. A few seek revenge." (Fisher 2016. 71.)

Before moving on, observe the dubious expectation in connection with non-arbitrary grounds above that what persons would prefer or accept to be loved for be taken into consideration. The more general question here is whether the whole "reasons of love" debate is normative or descriptive (empirical). If it is about why we *should* love others, it is presupposed that love can be willfully given a direction, for morality cannot demand the impossible. But even if we accept the reason-giving character of certain facts about the beloved (that they *should* be loved because they are blue-eyed and funny), it is rather doubtful that the beloved's preferences should figure in those reasons. (Below, I will follow de Sousa's descriptive approach in taking over results from empirical psychology as determining the scope of normative demands that can be placed on the potential lover.)

The quality of permanence is the odd one out among the characteristics attributed to love in this discourse. When attributing permanence to love, we stand on more shaky grounds than with exclusivity, irreplaceability, and loving "for oneself" – for "limerence", viz., love as we ordinarily understand it, is, while exclusive and intensively focuses on one person, temporally rather limited. Limerence, together with its time frame, will be addressed below.

To sum up the results so far, we love others 'for' particular properties as reasons, but not as separable from those persons. Rather, a complex set of individual properties may draw our attention to a person in a way that triggers the different

biologically-based psychological processes that will be described in Section IV. At least one further component of a plausible account is still missing, however: the historical (and, relatedly, narrative) dimension of love.

#### IV. HISTORICITY AND NARRATIVITY

The most conspicuous blindspot of the above theories is love's temporal character. Even if we love individuals on the basis of their qualities, they would seem to become sufficiently special to us not to be traded for others (with similar or different positive particular characteristics) during the course of a common history.

Nico Kolodny offered the following thought experiment to bring out the advantages of a history-based approach over a properties/reasons-based one. Why would a person lose his love for his wife due to amnesia? Is it because he has lost (grasp on) the reasons for loving her or because he has lost their common history? Imagine a non-fiction writer producing the biography of an admirable political activist based on thorough research, without personal acquaintance with his heroine. Years later they meet, fall in love, and get married. The biographer finds his wife to possess the very same qualities he had attributed to her without knowing her personally. Ten years later, he loses his later memories due to a medical condition, but he does remember the time he wrote the book. Will he continue loving his wife? Kolodny's view is that he will not, which allegedly demonstrates that it was not her qualities that made him love her in the first place (having perceived those qualities as exactly the same before and after meeting her) but their common history.

How plausible is this conclusion? Was the writer indeed dealing with the same properties or reasons in the two periods? On the basis of our conclusions above, the answer should be negative. The biographer could not have known those properties that made his future wife attractive to him at the time of writing the book. He was aware of certain properties *in abstracto*, but not in the way they were present in the individual. Personal acquaintance and history subsequently added concreteness to those qualities.

The connection between history and individual, non-abstract properties can also be approached from the perspective of the way in which the former shapes the latter. As Amelie Rorty remarks, love emerges on the basis of interactions between, and narratives involving, the subject and the object of love. In a love relationship that merits the name, both individuals, their attitudes and actions, are profoundly altered, at least temporarily. Thus, the properties of individuals, in addition to being concrete and trope-like in character, are further individualized as indexed to a particular relationship: one assumes particular qualities that other relationships would not be capable of providing in just that way.

Relationship histories are arranged into narratives by their subjects, individually and jointly, influenced by the sociocultural environment in which the relationship is formed. Part of the function of this narrative is to smooth over the potential rifts caused by the changes in persons' properties and events challenging the relationship. What de Sousa labels "founding lies" (self-hype, as it were) of a relationship may carry it over the rough patches. Thus, narratives provide a certain continuity. At the same time, they also individualize. The different narratives embedded in a relationship influence each other: elements of self-narratives feed into the joint relationship narratives, and self-narratives are also shaped by joint narratives, while both are under the impact of the paradigmatic narratives of the given culture accessed through channels like literature, movies, and social media. But what is the relationship between reasons and narratives?

#### V. REASONS, NARRATIVITY/HISTORICITY, AND BIOLOGY

Narratives may be constituted by (perceived) reasons. The content of some narratives are reason-giving properties, past or present, rather than events (e.g. 'I chose your mother because she was the prettiest girl in town'). Conversely, common history and narratives (e.g., the fact of, and narratives based on memories about, having spent 20 years together) may provide reasons to maintain a relationship and also perhaps to continue loving the other person.<sup>5</sup> In such cases, reasons and history/narratives are in line with each other. In multiple types of instances, this fails to be case. To be in a position to categorize those, and see the differences in the ways reasons may relate to the attitude of love, I'll use Helen Fisher's distinctions.

Helen Fisher, based on brain scan experiments performed on people in love, identified "three primary brain systems that guide mating and reproduction:" sexual drive; romantic attraction or "limerence;" and the feelings of deep attachment (Fisher 2016. 75; cf. Fisher 1998). For our purposes, the latter two are relevant, being focused on a single person. The second is what tends to be meant by romantic love: the condition characterized by focus on one person, the special significance of all that is attached them, intense sexual attraction, intrusive thoughts and vivid affective phenomenology (elation, hopes, anxiety, etc.). This condition, which might deplete resources and reduce functioning in other areas, typically lasts no more than 1.5-3 years according to brain studies: that is when dopamine and related neurochemicals start to decline.<sup>6</sup> At the same time,

<sup>5</sup> This may also be seen as a reason to *change*, rather than to stay in the relationship.

<sup>6</sup> This does not hold of a relatively small percentage of couples, as fMRI results and self-reports equally indicate. In these fortunate cases, intensity and sexual drive are maintained, the reward system is activated by the thought of the beloved even after decades of relationship,

attachment emerges, relying on oxytocin rather than dopamine, but also “emanat[ing] from the most primitive regions of the brain, near those that orchestrate thirst and hunger” (ibid, xii.) Attachment is associated with feelings of security, closeness, and reduced anxiety in the company of the partner.

Attraction and attachment, it seems to me, display different normative features. Attraction is based on properties such as “symmetry, the display of resources, the display of fertility, and/or other biological and behavioural factors that stimulate to whom one becomes attracted” (Fisher 1998. 30). It would not occur to us to censure someone for not being attracted to another person based on such properties. Also, justifying attraction only amounts to offering a subjective perspective on another person, without any commending value. There is no room for rational dispute in why a particular display of fertility by a certain person is more appealing to someone than another’s. That particular display (“trope,” as it was referred to above) is at the base-line of the explanation for the mating choice.

Love in the second, attachment sense, is most often associated with voluntarily imposed commitments, arising in the course of a long-term relationship. While it would seem much more sensible to cite reasons in this case than in that of limerence, commendation and censure by appeal to reasons are not so much about lack of the appropriate attitude as about the commitments and behavior associated with a long-term relationship (“Would you throw a 20-year relationship out the window?”). Inasmuch as maintaining an attachment, in the affective sense, can be achieved or supported by conscious effort, reasons can figure in the emotional side as well, however. Prolonged attachment can be helped by certain practices (Brubacher and Johnson 2017). When the long-term partner fails to exercise those practices that would sustain the attachment, persons could be held responsible for their own emotional distancing.

With the passing of time, keeping the relationship and attachment going may prove to be an uphill task. No doubt, the changing properties of the beloved, or new perspectives on those properties, can also have a role here. Having fallen in love with someone with a full head of hair, hourglass figure, or special sense of humour, it might be off-putting to find the hair or the figure go, or the humour turn out to be shared by four more persons in the same year in college. However, as the original attachment was not simply a matter of the hourglass figure or the sense of humour (many people having the same qualities in their own ways), falling out of love or switching will also be causally complex.

Here, the different factors described might work in tandem or be at odds with each other. The former case is less interesting (and perhaps more rare): the beloved’s properties, common history, and bio-psychological factors may carry the

---

while the anxiety of new love is much diminished (cf. Avcedo, Aron, Fisher, and Brown 2011).

attachment through the inevitable downturns. Let us turn to those cases where the three factors pull in different directions.

- (1) History vs. properties as reasons. People's external and psychological features, attitudes and behavior can change over time, sometimes drastically. New potential targets may beat old ones by a mile. Attachment hormones (and practical considerations) may only go so far to sustain the old relationship against the onslaught of such motives and considerations as the need for attention, novelty, etc., in combination with reason-giving properties of potential new partners.
- (2) History vs. biology. Limerence usually runs its course in 1.5–3 years. By that time, attachment, relationship, and mutual investment into that relationship solidify. Very crudely, dopamine and testosterone levels work against the relationship, oxytocin levels work in favour of it (Fisher and Thomson 2006). Here, personality types (dopamine vs. serotonin or oxytocin-driven) as well as attachment styles may play a role: those more dominated by oxytocin and having a secure attachment style are more likely to keep up a stable attachment.<sup>7</sup>
- (3) Reason-giving properties and history vs. biology. Partners against whom violence has been perpetrated often choose to remain in the abusive relationship. Apart from (social or economic) pressures, the reason tends to be found in bio-psychological factors, such as attachment and co-dependence, which may outweigh reasons emerging from the history of the relationship, personality traits and behavior of the aggressive partner.
- (4) Reason-giving properties vs. narratives. With plenty of reasons for abandoning a failed attached relationship, one might be held captive by its "founding lies." (Such narratives may also complement the bio-psychological factors in maintaining an abusive relationship.) Illusions of a special union and the super-power of overcoming any hardships together may trump sombre realities.

## VI. CONCLUSION

This paper has argued against the prevalent framing of the "reasons of love" problem in terms of the "reason-based" and "no-reason" views, even when complemented by an historical dimension. Observing the non-exhaustive character of these two theoretical possibilities and the ambiguity between descriptive and normative formulations, we have noted the tendency, already present in Plato and also shaping the present debate, of conceiving properties as potentially imposing themselves on the would-be lover, without their agency, psychological

<sup>7</sup> For a description of different attachment styles, see Brennan and Shaver 1995.

makeup and biological circumstances assigned a more substantive role. Here, an attempt has been made to acknowledge some of the intricacies related to these factors, and their interplay, in the emergence of the attitude of romantic love.

## REFERENCES

- Avecedo, Bianca P. – Arthur Aron – Helen E. Fisher – Lucy L. Brown 2012. Neural Correlates of Long-Term Intense Romantic Love. *Social Cognition and Affective Neuroscience* 7/2. 145–159.
- Brennan, Kelly A. – Phillip R. Shaver 1995. Dimensions of Adult Attachment: An Integrative Overview. In Jeffrey A. Simpson – W. Steven Rholes (eds) *Attachment Theory and Close Relationships*. New York, Guilford Press. 46–76.
- Brubacher, Lorrie L. – Sue Johnson 2017. Romantic Love as an Attachment Process: Shaping Secure Bonds. In Jennifer Fitzgerald (ed.) *Foundations for Couples' Therapy: Research for the Real World*. New York, Routledge. 8–19.
- Campbell, Sue 1998. *Interpreting the Personal: Expression and the Formation of Feeling*. Ithaca, Cornell University Press.
- Dancy, Jonathan 1983. Ethical Particularism and Morally Relevant Properties. *Mind* 92. 530–47.
- de Sousa, Ronald 2015. *Love: A Very Short Introduction*. Oxford University Press.
- de Sousa, Ronald 2016. Love and Reason: Reflections on Themes from Peter Goldie. In Julian Dodd (ed.) *Art, Mind, and Narrative: Themes from the Work of Peter Goldie*. Oxford, Oxford University Press.
- Fisher, Helen 2016. *Anatomy of Love: A Natural History of Mating, Marriage, and Why We Stray* (revised and updated). W.W Norton and Company.
- Fisher, Helen 1998. Lust, Attraction, and Attachment in Mammalian Reproduction. *Human Nature*. 9/1. 23–52.
- Fisher, Helen – J. Anderson Thomson 2006. Lust, Romance, Attachment: Do the Side Effects of Serotonin-Enhancing Antidepressants Jeopardize Romantic Love, Marriage, and Fertility? In Steven M. Platek – Julian Paul Keenan – Todd K. Shackelford (eds.) *Evolutionary Cognitive Neuroscience*. Cambridge/MA, MIT Press.
- Frankfurt, Harry G. *The Reasons of Love*. Princeton University Press, 2004.
- Kivy, Peter 1999. Feeling the Musical Emotions. *British Journal of Aesthetics*. 39. 1–13.
- Kolodny, Niko 2003. Love as Valuing a Relationship. *The Philosophical Review*. 112/2. 135–189.
- Plato *The Symposium*. 2008. Trans. by M. C. Howatson, edited by Frisbee C. C. Sheffield. Cambridge University Press.
- Scanlon, Thomas 1998. *What We Owe to Each Other*. Belknap Press of Harvard University Press.
- Smuts, Aaron 2014a. Normative Reasons of Love I. *Philosophy Compass*. 9/8. 507–517.
- Smuts, Aaron 2014b. Normative Reasons of Love II. *Philosophy Compass*. 9/8. 518–526.

DAVID WEBERMAN

## What is an Existential Emotion?

My aim in this paper is to make more precise the idea of an existential emotion. I want to explain exactly what it might mean to say that there is a subset of emotions which qualify as existential. The framework for my analysis follows Heidegger's account in *Being and Time*.<sup>1</sup> While I follow that account, I will be adjusting the vocabulary, probing it in certain ways that he himself does not and building on it to reach some new insights. In fact, my central question about which emotions are existential and what makes them so, uses the term "existential" differently from Heidegger to pick out a certain character, potent and insightful, that some emotions can have.

The paper has three parts: In the first part, I will need to say something about my use of the concept of emotion (a term which is absent from Heidegger's text). Despite what some commentators on Heidegger and other theorists of emotion say, I will argue that Heidegger's notion of *Befindlichkeit* is essentially about we call emotions and that emotions come in two types: i) moods and ii) object-specific emotions. I will argue that Heidegger takes both types (correctly) to be intentional, that is, directed at or about something. This something is their "object" in a phenomenological sense of that term. I will say what the two types, moods and object-specific emotions, have in common and what sets them apart. In the second part of the paper, I want to use the notion of existential in a way that applies to certain emotions (It might apply to other things as well). I will then ask which emotions can be existential and what makes them so. Is it only moods that are existential? And, among moods, are there certain of them such as *Angst* that have a special claim to being existential in the sense used here? This will lead to me to the third part of the paper in which I pursue the various ways in which emotions can be seen as existential depending on how and what they disclose. In the end, I will pres-

<sup>1</sup> Heidegger 1962. Page numbers in the article are indicated by SZ and the German pagination included in the margins of the English edition.



ent a sort of template for the existential character of emotions and suggest how the notion of existential might be understood to apply to other things besides emotions (such as ideas or literary works).

### I. BEFINDLICHKEIT, EMOTIONS AND MOODS

Heidegger's neologisms usually have an important point, but they can complicate matters, especially if they only allow us to speak in his way and not to correlate what he says with claims made in our ordinary manner of speaking. In a short but important part of *Being and Time*, Heidegger discusses the notion of *Befindlichkeit* as one of three ways (*Weisen*) constitutive of the thereness (the "Da") of Dasein. Most of his discussion is about moods (*Stimmungen*). But he also discusses fear which he calls not a mood, but a mode (*Modus*) of *Befindlichkeit*. But what is a mode of *Befindlichkeit*? And what is *Befindlichkeit* given that it encompasses both moods and modes which are not moods?

My suggestion is the following and it does indeed conflict with many commentaries on Heidegger as well as with some intuitions about how to use the English word "emotion." First, Heidegger's modes of *Befindlichkeit*, such as fear (or anger or jealousy) refer to what we ordinarily call emotions. His analysis of fear makes this quite plain. Second, we can take Heidegger's moods to denote roughly what ordinary speakers mean by "moods", though he fleshes out the notion of a mood in a quite distinctive manner to which I'll return shortly. Third, if moods and object-specific emotions belong to one single category, how shall we understand that category? What is it that Heidegger's *Befindlichkeit* encompasses? At this point, I would like to make a controversial move. We can accept what much of what Heidegger says about *Befindlichkeit* – that it is a way of finding oneself, that it is a kind of attunement, that it is a condition for the possibility of anything mattering to us which is a condition for the possibility of anything meaning anything determinate to us – yet also hold that *Befindlichkeit* covers the entire domain of human emotions.

This claim requires a brief defense because it conflicts with the way that many people in English and other languages use the term "emotion." Many are inclined to think that moods are not emotions because emotions are always object-specific, while moods are not. But must or should we use the term "emotion" in this way? Doing so might prevent us from seeing that moods and object-specific emotions have something in common, something that is hard to define but gets at how we feel about things in a, well, emotional way. This is strongly indicated by the fact that anger or sadness can be both an object-specific emotion and a mood. This can't be a mere coincidence or a mere linguistic oddity. Both are feelings of a certain sort. Not any type of feeling. They are not, for instance, like feelings of heat or pain which are what philosophers call

sensations. But a certain type of feeling that seems best captured by the term “emotion.”

Heidegger doesn't use the term emotion at all. Not only is it less common in German than in other languages, its etymology contradicts one of Heidegger's core commitments. Etymologically, it comes from the notion of moving out of some perhaps neutral state. But Heidegger thinks that we are always in some mood and that there is no neutral, mood-less state. And so it's not clear that the etymology helps much here or that it should limit us in any way. We are free to use words in ways that depart from their origin. And we might need an umbrella term for both moods and object-specific emotions. Although Heidegger uses the term “feelings” (*Gefühle*) at one point, this won't do the job because, as mentioned, it is too broad in that includes sensations such as warmth and pain. At another point, Heidegger contrasts his account with earlier philosophical theories of the “affects.” Yet the term “affect” is clinical or academic and possibly misleading since, in medicine, it highlights the largely bodily or facial expression of feeling, not the feeling itself. So, for lack of a better term and for the sake of convenience, I will use the term “emotion.” But much of what I will say does not depend on my choosing that term, except insofar as it presupposes that moods and object-specific emotions belong to a single category.

Now, let me return to moods. If it is a type of emotion, what type is it? Because it contrasts with object-specific emotions, it is natural to think that moods, being non-object-specific, are diffuse or generalized emotions. Yet some philosophers take issue with this way of seeing things. For example, in an essay on Heideggerian moods, Matthew Ratcliffe writes:

It is commonplace to regard moods as generalized emotions, meaning emotional states that are directed at a wide range of objects [...]. A mood, for Heidegger, does not add emotional color to pre-given objects of experience [...]. [A] a mood is not a generalized emotion. It is not a way in which any number of entities appear but a condition of entities being accessible to us at all. (Ratcliffe 2013. 159.)

Ratcliffe is right to say that, for Heidegger, having a mood is an enabling condition. It enables our access to entities by allowing them to matter to us and thus allowing them to mean something in particular to us. Moods, for Heidegger, do not merely “color” objects that are already accessible because already individuated and fixed with a prior determinate meaning. But none of this inconsistent with moods having a generalized directionality. Having some mood or other may make possible the accessibility of entities, while at the same time it is also the case that the moods we have are generalized such that they are directed at not just this or that thing, but anything that comes its way. Moods are both constitutive (if Heidegger is right) *and* generalized background emotions that “cloak” whatever we encounter in an object-unspecific way. Note the word “cloak”

here. I use it, despite its similarity to the word “color” because I want to hold on to the generalized character of moods without giving the impression that the role of moods is at all secondary or superficial.

Now, if moods are generalized emotions, then it would seem to be that when I’m in a sad mood, I’m sad about everything and that the *object* of my sad mood is, as it were, everything. But I do not want to say this. I want to say, following Heidegger as we shall see that moods cloak everything or anything but that the object of our moods is actually something else. Before I say what it is, let me first say something about the idea that moods like other emotions have objects because all states of consciousness, if I can use that non-Heideggerian parlance here, have objects and moods are one type of conscious state.

Do moods and the modes of *Befindlichkeit* such as anger have *objects* for Heidegger? Heidegger does not talk about the “objects” of *Befindlichkeit*. Heidegger does not want to use, of course, the word “object” (“*Objekt*”, or even “*Gegenstand*”) because it implicates what he takes to be an untenable dualism of a self-contained subject and a subject-independent object. He speaks of *Dasein* and for physical object he uses the term “*innerweltliches Seiendes*” for such things as tables and chair. But the philosophical term “object” does not always refer to physical objects; it sometimes refers to what are called intentional objects, that which conscious states or acts are about or directed at.<sup>2</sup> (I’ll set aside for now Heidegger’s avoidance of the term “consciousness.”) But Heidegger has another term for what emotions are about or directed at. It denotes, in this context, exactly what the term “intentional object” denotes. That term in Heidegger is “*Wovor*.” In his discussion of fear as a mode of *Befindlichkeit*, he says that such modes have three aspects: the “*wovor*” (the in-the-face-of which) of fear, fearing itself, and the “*worum*” of fear (“that about which or for the sake of which we fear.”) The fearing or emoting itself is the experience of being in a particular state (e.g., fearing rather than loving). When we fear an approaching bear, the bear is the “*wovor*” and the “*worum*” is always *Dasein* itself, its survival or well-being (regardless of whether it is mine or someone else’s). Moreover, Heidegger says in this passage that the tripartite structure of emoting itself, the “*wovor*” and the “*worum*” applies not only to modes such as fear but also to *Befindlichkeit* generally (SZ 140). He later applies this same three-part structure to *Angst*. So the point is that all emotions, including all moods, have a “*wovor*” or, as

<sup>2</sup> See Tim Crane 2008. 489, on the idea of an intentional object. Crane discusses briefly whether moods have objects though he does not reach a conclusion. Analytic arguments for propositional objects should not be regarded as unacceptable to Heideggereans. First, this idea has its roots in Brentano and Husserl. Second, while Heidegger rejects talk of “subjects” and “consciousness” it is still the case that *Dasein* (unlike a stone) is something to which things are disclosed and thus is something like a subject and has something like consciousness. (But this is of course denied by more radical readers of Heidegger and by Heidegger himself.)

I'll call it here in light of contemporary philosophy of mind, an object. (Note, by the way, that the object or "*wovvor*" of an emotion is not necessarily the same as its cause. For example, I may be nervous about an interview which is the *object* of my nervousness even if its cause is too much coffee.

So, now to return to the question of the object of moods: If moods have objects, i.e., something that they are directed at, and if moods are not object-specific but generalized, then it would seem that the object of a mood, such as a sad mood, is everything. But this is, I think, not quite right. It is plausible to hold that everything can be an object of *thought* or *belief*. But it is implausible that everything is the object of our moods, at least, that is of ordinary moods (I'll come back to the distinction between ordinary and a class of special moods later in the paper.) If we believe that everything is physical or, alternatively, that everything is created by God, then our mind is directed at a certain "object," namely, everything, at least in a certain aspect. But moods would seem to be different. Everything is not the object of a sad or angry mood because everything takes in far too much. Is one really angry or sad about everything, about every single thing such that it includes everything *down to the very last thing*? This seems unlikely. One can have a belief about everything because believing something can come in one fell swoop, but I doubt that one can be sad about every last thing all at once. It would be more correct to say that a sad mood is not about everything but about anything, that is, anything that comes my way. It cloaks or casts its pall on whatever I happen to encounter. This is its generalized character, anything not everything. Yet note that "anything" is a variable, meaning that it has the logical form "For any x, if x comes my way, x will be seen as sad or sadness-evoking." But it seems odd to think that the object of my mood has a form involving this kind of variable. A variable seems to be an unlikely object of my moods. So, it seems reasonable to think that the object of a mood must be something else. In fact, this is Heidegger's view. Everything is not the object of a mood, rather there is something else that is. It strikes me that Heidegger has a view about what that object is and it strikes me as a rather good answer to our problem.

For Heidegger, the object of a mood is neither a particular object or state of affairs *nor* everything *nor* the variable "anything"; it is something very particular. Moods bring us up against the fact that Dasein is delivered over ("*überantwortet*") to being and consequently that Dasein is an entity that "must be existingly" (*dass existierend zu sein hat*) (SZ 134). More briefly, moods are directed at the bare fact that Dasein "is and must be", "*dass es ist und zu sein hat*" (SZ 134). Heidegger also formulates this point by saying that we are "thrown", i.e. that we find ourselves existing (and existing in particular circumstances) without having chosen to do so. What we are thrown into is not just that we must deal with having to exist in a generic sense of "exist," but that we must exist in the specifically Heideggerean sense of "exist" replete with all of the necessary and universal

features (*Existenzialia*) that are constitutive of Dasein, e.g. being *Mitdasein* or social, being mortal or *Sein-zum-Tod*, being a project, etc., etc. In other words, our moods have as their object our having to be being-in-the-world. When I'm in a sad mood, I'm sad about that and when I'm in an irritated or a happy mood, I'm irritated or happy about having to be existingly, having to be being-in-the-world. I may not be conscious that being-in-the-world is the object of my mood, since moods are not transparent in their structure, but that is what all moods, on this account, are directed at.

How can we be so sure that being-in-the-world is the object of our moods rather than everything? Can we test that claim for its plausibility? It seems that we cannot test it by asking ourselves what we're consciously sad about when we're in a sad mood because the object of a mood is not always conscious. Introspection does not reliably turn up the object of consciousness. We can however ask ourselves whether we're really sad about cups and saucers and chairs and tables which are part of everything. The clear answer is that we're not sad about cups and saucers and chairs and tables. So we're not sad about everything. By default, then it must be something else. I would propose, whether we know it or not, is our having to be existingly our having to carry on under current circumstances is the better answer. At the same time that the object of our moods is being-in-the-world, our moods cloak anything that comes their way which is compatible with these things first being made available through the having moods, being somehow affect, to begin with.

To conclude this section, then: My reading of Heidegger is that *Befindlichkeit* is a fundamental and necessary aspect of our existence that picks out the emotional side of our existence. These emotions come in two kinds: object-specific and generalized moods. Both kinds of emotions have objects. The objects of moods is our being-in-the-world or, more precisely, our having to be being-in-the-world, which is to say our having to exist with all that's built into Dasein's existence and all the givens of the existence of any particular Dasein. (This having-to-be need not elicit sadness or despair, it may be encountered in delight, when we're in a very good mood or equanimity, when we're in a more neutral, serene mood.)

## II. THE EXISTENTIALITY OF EMOTIONS

Might it make sense to say that some emotions are existential? Are the emotions that are existential moods rather than object-specific emotions? Or all or only some moods existential. Here I want to use the term "existential" not so much in Heidegger's technical sense but in a sense more familiar to us from a more generic sense of the term that happens to bear the mark of influence from existential philosophy. I have in mind an adjectival use of that term such that it can

describe certain phenomena such as emotions but other things as well, such as artworks or experiences. Heidegger hardly, if at all, uses the term “existential” in this way. His adjectival use is to designate certain structures as existential if they are necessary, universal and constitutive features of being Dasein. But I think there is value in using the term in another way if it brings out a certain aspect of things of real interest to us, namely an aspect that spotlights the human condition and that contrasts with our everyday absorption in very particular and often very parochial concerns and projects. Let me define it here as follows: something is existential if it brings to light in a profound manner something central to the human condition.

Before proposing my own idea of what makes an emotion existential, consider the following idea provided by Matthew Ratcliffe, the philosopher quoted above, who has developed the idea of an existential feeling in a book titled *Feelings of Being*. Ratcliffe argues for calling them “feelings” rather than emotions, moods or affects for interesting reasons that I won’t go into here. What is important here is his use of the notion that such emotions or feelings are *existential*. He writes:

Existential feelings are both ‘feelings of the body’ and ‘ways of finding oneself in a world’. By a ‘way of finding oneself in the world,’ I mean a sense of the reality of self and of world which is inextricable from a changeable feeling of relatedness between body and world. (Ratcliffe 2008. 2.)

I will not go into the “bodily” aspect which is unfortunately rather neglected by Heidegger. I want to focus on Ratcliffe’s notion that existential feelings are about “a sense of the reality of self and of world” which, as he goes on to say, is about “our relatedness to the world which can range from a feeling of belonging to the world to a feeling of detachment or alienation. Ratcliffe says that this detachment can manifest itself in a range of feelings such that reality can seem “surreal, unfamiliar, uneasy, not quite right or too real” (Ratcliffe 2008. 3). This is not a propositional attitude or belief, Ratcliffe says, but a feeling, i.e. something felt, felt in the body but about something outside of the body, namely and in a word, reality.

To my mind, Ratcliffe has correctly identified and nicely described an important phenomenon, one that lies at the heart of much existential philosophy, from Kierkegaard onwards, and much existential literature (Kafka, Camus, Beckett to cite just a few examples). It also fits and illuminates various pathologies he discusses such as depression, schizophrenia, etc. But his characterization of “existential” in this way seems to me too narrow. It seems to be organized around whether we are healthy, at ease and connected or whether we suffer from some pathology of disconnection. In plotting mood along a single axis of connectedness and detachment, it overlooks that there can be and often is much more to

existence than whether or not we feel at home in it. Let me turn now back to Heidegger for a wider way of characterizing what might make emotions and other things existential in the sense at issue here.

As we have seen, moods have as their object our having to exist in the manner distinctive of Dasein. This would be a broader answer than Ratcliffe's because existence has so much built into it, from *Mitdasein* (being a social creature), to being a project, to being mortal, to having a tendency toward inauthenticity, etc. Does this mean all moods are existential in bringing us up against the fact of having to exist? At this point, it is important to turn to Heidegger's discussion of *Angst*. It gives us a more specific sense of the object of moods – all moods, as well as special ones, the most prominent of which is *Angst*. It will also provide a distinction that can be marshalled for seeing what existential in the sense here means and why certain emotions have a special claim to being existential in that sense.

In his crucial section on *Angst*, Heidegger says that *Angst* is a fundamental and distinctive type of *Befindlichkeit* (“*Grundbefindlichkeit*”, “*ausgezeichnete Befindlichkeit*”, SZ 182, 184). In *Being and Time*, he abstains from calling it a mood. In his later “What is Metaphysics,” he does indeed refer to it a mood (Heidegger 1929/1993). I think the reason for this is while most moods are typically unnoticed backgrounds to experience, the experience of *Angst* is felt when one is in it. In fact, it is so dominant, so overwhelming that all else falls away. Still, it is a mood because it is an emotion or feeling and it is one that has a generalized character cloaking anything (and, even as an exception to the rule, everything, as we will see). It is, in a nutshell, an intense experience of homelessness, detachment, uprootedness.

What is the object of *Angst*? Heidegger says it is “being-in-the-world as such” (“*Das Wo vor der Angst ist das In-der-Welt-sein als solches*”, SZ 186). His description of the object of *Angst* is more concise than his description, earlier in the text, of the object of moods in general. However, it seems clear that both moods generally and *Angst* in particular have as their object being-in-the-world, i.e. having to be being-in-the-world. Yet Heidegger adds something to the object of anxiety, namely, the “as such.” While the object of *Angst* is, like the object of moods, being-in-the-world, only *Angst*, unlike other moods, has as its object being-in-the-world as such. What does the “as such” add? I would suggest that the “as such” abstract away from a particular individual's having to be being-in-the-world. Garden-variety moods, such as irritation or sadness, run up against and disclose my *particular* having to exist in a *particular* set of circumstances at a given point in time. *Angst* runs up against and discloses what it is for any Dasein to have to be Being-in-the World in any set of circumstances at any time. In *Angst*, entities fall away as unimportant (“*ohne Belang*”) and as insignificant. This means that we come to see a certain philosophical truth that the world is, otherwise, except in the throes of *Angst* a network of significance and that Dasein's existence consists



in care, as specified by the Existentialia. My point is this: i) all moods have as their object being-in-the-world; ii) garden-variety moods have as their object my particular having to be being-in-the-world in particular circumstances; iii) Angst has at its object anyone's having to be being-in-the-world. That is, the object of Angst, is being-in-the-world as such. It imparts a certain insight into the human condition.

There are a few more things to say here. First: Is Angst the only special mood, while all other moods are garden-variety? If we continue to follow Heidegger (which we needn't do, of course), it is not the only special mood. In his essay *What is Metaphysics* (1929), he suggests that there may be other moods that are like *Angst* such as a certain kind of love, or a certain kind of joy, or a certain kind of boredom. In fact, in his 1929–1930 lecture *The Fundamental Concepts of Metaphysics: World, Finitude, Solitude* (Heidegger 1995), he develops the idea of deep boredom, a mood which is special because it reveals the nature of what it is to be *Dasein* and what it is to have a world. Deep boredom may cloak everything, but its object is being-in-the-world, not my being-in-the-world but anyone's and everyone's being-in-the-world. This is the point of adding that being-in-the-world is revealed as such.

I will not follow up on other possible special moods here and why it is that they qualify as special, that is, how it is that they reveal what it is, in general, to be *Dasein*. That would take another essay. What I want to say here is that these special moods are excellent candidates for claim of being existential because they show us something about the general nature of existing in a way that garden-variety moods do not. *Angst*, and a fortiori, any mood is existential insofar as and because its object is existing as such.

Heidegger says something else about *Angst* that is relevant here. Earlier I said that moods cloak anything, though not exactly everything. But Heidegger says that the special moods such as *Angst* or deep boredom reveal things as a whole (i.e. everything) and also that they reveal (the) nothing (*“das Nichts”*). Thought, according to Heidegger, gives us only a formal idea of these, the sum-total of all that is and its negation. But, according to Heidegger, special moods and only special moods reveal the real thing: not merely the idea of everything or the idea of nothing, but everything or nothing itself. This claim goes beyond what has been said until now. I am not convinced that it is right. I think the position is that there is a difference between the formal idea of everything and an everything we can encounter that is the real McCoy (the genuine article). The same goes for nothing. There is the formal idea of nothing, that we can think, and then there is the genuine article, nothing itself as encountered in *Angst*. There are two questionable moves here. One is that there is a distinction that can be upheld in the two cases and the second is that we have access to everything and nothing in some non-idea-like, non-formal, non-propositional form. Whether this is defensible or not, it is an extra move that is not required by



the argument above. If it is defensible, then it would mean that we could call the special moods, the “as such” moods, totalizing moods because of their special contact with everything and nothing.

Back to the term “existential.” Let me add here a distinction between two levels of something being existential in the present sense. We can summarize our results stating the following: Garden-variety moods are *minimally existential* because their object is being-in-the-world but only insofar as it concerns my particular being-in-the-world. Special moods are *maximally existential* because they run up against and disclose and have as their object being-in-the-world as such. According to Heidegger some moods (what I have called special and maximally existential moods) are totalizing because they reveal everything and/or nothing. One might go on to say that maximally existential moods are philosophical because they reveal something about the general character of being human. Finally, I would suggest that one can import this notion of existential to artworks and other things and experiences. What I mean is that a novel, poem or play is minimally existential to the extent that it reveals the being-in-the-world of a particular Dasein in particular circumstances and an artwork is maximally existential to the extent that it reveals being-in-the-world as such or what has often been called the human condition.

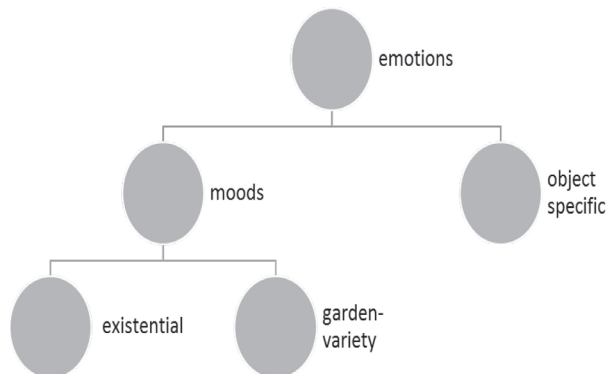
### III. RETURN TO OBJECT-SPECIFIC EMOTIONS

My argument has led to this point: While moods are existential, either minimally or maximally, object-specific emotions are not at all existential. Object-specific emotions are object-specific so they may well be of great importance, in some cases about life and death issues, they don’t tell us about being-in-the-world so they are not existential. They are about, have as their objects dangerous animals (and even harmless spiders) or rude car drivers or, to be more positive, caring parents or one’s favorite football team. Heidegger’s own treatment of fear suggests that object-specific emotions are not directed at our being-in-the-world. While there “*worum*” is always Dasein, their “*wovor*” is never anything but intramundane.

But having thought a bit more about the matter, it strikes me that this might be wrong or far too hasty. Some object-specific emotions might well have an existential component or at least an existential follow-up. A couple of examples: A friend of mine recently felt deep grief when her beloved cat died and it threw her into what could legitimately we called an existential crisis – a crisis about the nature of being-in-the-world and our aloneness underneath it all. It brought out for her, what Heidegger might call, the reality of a certain deficient mode of *Mitdasein*. Similarly, being upset about the sudden diagnosis of a life-threatening ailment might precipitate a kind of existential crisis in which all sort of

questions about the nature, value and point of being-in-the-world come to the fore. Or, modifying an example from Schopenhauer, a storm at sea might elicit object-specific emotions that border on or lead to existential ones. In such cases, the object-specific emotion seems to precede both temporally and perhaps even logically the existential mood or existential experience. So, one might say that it is not the object-specific emotion that is existential but a certain mood that it triggers. Now, Heidegger, to my knowledge, does not see or mention this. In the closing paragraphs of the section on fear, he discusses how fear, if it is sudden, can slip into *Erschrecken*, or into *Entsetzen*, if it is combined with *Grauen*. But none of these states is existential in the sense meant here, none are about being-in-the-world. Heidegger's section on object-specific emotions is very short. But I would propose the following: while moods are directly existential, minimally or maximally because their object is being-in-the-world, object-specific emotions can be indirectly existential, because while their object is not being-in-the-world, they can very quickly turn into an experience which does indeed have being-in-the-world as its object. Grief for one's dead cat can in principle bring about a state that raises existential flags.

To summarize this interpretation, I provide the following diagram:



#### IV. RUNNING UP AGAINST, DISCLOSING AND REVEALING

The reader may have noticed the following terminology used in my analysis. I have said that moods have objects that they “run up against,” or “disclose” or “reveal.” In Heidegger’s text, there is also use of various descriptors such as “*begegnen*,” “*aufbrechen*,” “*sich zeigen*,” “*enthüllen*” and most prominently “*erschliessen*” (disclose). There is a sentence in Heidegger’s text that has especially tipped me off to what I would call a certain ambiguity in these words and perhaps in the all-important Heideggerean notion of disclosure (or disclose and disclosive). On SZ 185, Heidegger introduces Angst as a “methodologically disclosive” (“*metho-*

*disch erschliessende*”) *Befindlichkeit*. But, earlier he had said that all *Befindlichkeit*, all moods and emotions are disclosive. So why does he say here that *Angst* is disclosive and methodologically so. Well, perhaps his point is that while all emotions are disclosive, only *Angst* is methodologically disclosive because it helps us to the philosophical insight of what makes the world a world and what makes *Dasein* care. In fact, I think this is so. And it supports the distinction I have made between garden-variety and special moods, where the former are about my being in the world and the latter about being in the world as such which has a certain methodological privilege because it tells us about the philosophical nature of *Dasein* and world. This point signals, I think, a certain ambiguity in the term “disclose.” When moods disclose the world as irritating or sad or worthy of joy, they do so by letting things appear and matter in a certain way. But they don’t necessarily reveal anything new or give us any new insights beyond just the world appearing in a certain way. But when *Angst* is experienced, we are led to new insights. New things are revealed to us that go beyond things appearing in a certain way. It seems that one can distinguish two senses of disclosure: i) letting things appear in a certain way and ii) giving us new revelations or insights. This may even be in line with a criticism levelled by Tugendhat (1967) about Heidegger’s theory of truth as disclosure and unconcealment. That is, truth as disclosure only tells us that what truth presupposes, namely, letting things appear in a certain way. It is another matter altogether to say whether things appearing in a certain way gives us any purchase on the world and whether it gives us a new insight that was absent until a particular disclosure took place. Perhaps this is another way in which garden-variety moods are different from special moods. While garden-variety moods disclose in the first way, letting things appear in a certain way, only special moods give us new insight or realizations. These new realizations are maximally existential, giving us new insights into the “as such” of our existence. In fact, in this sense, special moods are not only maximally existential, they are philosophical. This is a reason to think that philosophy does not rely on arguments alone. Certain special moods might have a role to play in advancing philosophical enlightenment about the human condition.

## REFERENCES

- Crane, Tim 2008. Intentionalism. In Brian P. McLaughlin et. al. (eds.) *The Oxford Handbook of Philosophy of Mind*. Oxford, Oxford University Press. 474–493.
- Heidegger, Martin, 1927/1962. *Being and Time*. Translated by John Macquarrie and Edward Robinson. Oxford, Basil Blackwell.
- Heidegger, Martin 1929/1993. What is Metaphysics? In David Farrell Krell (ed.) *Martin Heidegger: Basic Writings*. Revised and expanded edition. London, Routledge. 93–110.

- Heidegger, Martin 1995 (lecture from 1929–1930). *The Fundamental Concepts of Metaphysics: World, Finitude, Solitude*. Bloomington, Indiana University Press.
- Ratcliffe, Matthew 2008. *Feelings of Being: Phenomenology, Psychiatry, and the Sense of Reality*. Oxford, Oxford University Press.
- Ratcliffe, Matthew 2013. Why Mood Matters. In Mark Wrathall (ed.) *The Cambridge Companion to Heidegger's Being and Time*. Cambridge, Cambridge University Press. 157–176.
- Tugendhat, Ernst 1967. *Der Wahrheitsbegriff bei Husserl und Heidegger*. Berlin, De Gruyter.

LORE HÜHN

## Com-compassion: On the Foundations of Moral Philosophy for J. J. Rousseau and Arthur Schopenhauer\*

There are impressions to which one only needs enough unprotected exposure to know that they ought not to be. It is generally our emotions which inform us about this judgement (Theunissen 1983. 41–44). There is hardly a phenomenon which brings this before our eyes with more urgency than being affectively impacted by the suffering of others. As everyone knows from personal experience, this state of being impacted belongs to the form of reactive affects which can scarcely be translated into objective statements. This form of affectivity furthermore refers us back to ourselves, not directly and without mediation, but in a way which is infiltrated through and through with the pain of the other. It confronts us with ourselves, although this self-confrontation has already traversed the way of the other without seriously posing the question as to how this person here has come to enter into a relationship with that person there in the first place (Waldenfels 2000. 290). Compassion – with which the following is concerned – is mediated through an experience of the self which strikes the subject in his innermost core and, for Schopenhauer, throughout the entirety of its existential self-realisation. It concerns the subject's fragility and vulnerability – a vulnerability which numbers among the basic elements of our finitude, and thus ultimately of life itself (cf. Hühn 2007).

### I. THE ETHICS OF COMPASSION AS AN ALTERNATIVE MODEL TO AN ETHICS OF PERSONAL RECOGNITION

Along with the phenomenon of being affectively impacted by the suffering of others, a relation which is always already moving between self-relation and relation to the other is on the table. Yet neither side of this relation could be

\* Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2193/1 – 390951807. I thank James Fisher very much for the translation of the article into English.

separated from the other precisely because it has been determined long before any reflection that they are mediated through one another (Das 1995); indeed, as Schopenhauer suggests, they may even form a unity. It is no coincidence that Schopenhauer provides plausible proof of an essential identity of all living creatures founded on a metaphysics of the Will by means of explicit reference to the phenomenon of compassion. In a clear and deliberate dissociation from a morality based on theories of recognition, Schopenhauer anchors this identity so deeply in the fundament of our practical relation to ourselves and the world that we derive our right to recognition from this identity. These rights extend to all other creatures which are capable of suffering, animals included. The avowed intention of the ethics of compassion is to prevent a gulf before it happens, specifically that gulf which emerges between self and other and which should be overcome because of this separation. The ethics of compassion consciously opposes itself to the Kantian-idealist model of personal recognition by ensuring that all dualistic models are defanged before the separation of self and other first appears through reference to the monism of a hidden initial identity of all beings capable of suffering. Schopenhauer's ethics of compassion has a further anti-Kantian point in that it extends the obligatory character of moral norms to nature by valorizing a single feeling, namely that of compassion. It thus reaches a domain which is excluded from the traditional ethics of purely personal recognition. In a way which upends the anthropocentric orientation of the Cartesian paradigm, the feeling of compassion extends the frontiers of subjectivity. Yet it does not experience the conditions which render it possible, namely the totality of a nature founded on a theory of Will and including humans as well as animals.

## II. ETHICS OF COMPASSION AND UTILITARIANISM: THE CRITERION OF THE CAPACITY TO SUFFER

From an ethical perspective, being moved by the suffering of others is not uncommonly connected with the practical imperative to intervene in a situation and reduce the suffering where possible (Maio 2006). It is no coincidence that the moral obligation to eliminate suffering belongs to the basic components of all ethics which orient themselves solely according to the feelings of those to whom we have obliged ourselves morally. Utilitarianism and the ethics of compassion, as much as they differ in other aspects, have this in common: they emphasize the capacity to suffer as a criterion for morality in such a way that the capacity to suffer entails the obligation not simply to capitulate impotently before the suffering of others, but rather to prevent or reduce it wherever and however it occurs (Tugendhat 1997). Doubtless, this ethical attitude goes much further than an attachment to the immediate horizon of concrete observation

(Birnbacher 2006). This is readily apparent in the case of utilitarianism, with its criteria of sustainable efficiency for measures serving to prevent suffering; in the case of the ethics of compassion, it has yet to be discovered.

In order to counter misunderstandings, it must be emphasized in view of Schopenhauer's ethics of compassion that it by no means aspires to the strict expansion of a concrete and spontaneous act of compassion according to the standards of situational ethics which intervenes in certain occasions of need; nor is it a naive pantheistic feeling of unity which extends to each and every thing in the manner of a one-dimensional anthropomorphism. The ethics of compassion is much more concerned with a readiness to recognise others not as specific persons, but as sufferers and only as such – a readiness which extends beyond the immediate situation. This means perceiving this or that person with a conscious abstraction from all the other determinations which otherwise constitute her existence. To be sure, compassion begins with the familiar suffering of others which I encounter at close proximity and develops generically from there. Yet an emotionally oriented state of being impacted, dependent on a situation and bound to perception, doesn't go very far, particularly in the case of an acute given pain, and certainly not far enough to serve as a foundation for a morality which would be deserving of the name. Such a morality is not episodic; it does not decide from case to case and does not permit any exceptions but extends with universal validity to all – all that we count among the universe of moral addressees. There is also no doubt that being affectively impacted through suffering does not replace moral judgements. And yet if the impact is missing, then the elementary experience of values which precedes every moral judgement is generally missing as well (Spaemann 2001). It is in particular the practical impulse needed for moral action which is missing. This impulse leads us to do that which, after considering all foreseeable results, will sustainably limit suffering and introduce measures to prevent it. Horizons of distance and proximity; behavior which is personally addressed and yet frees itself from personal relations: these fall into a paradoxical connection which Walter Schulz recalls again and again because to his eyes, it continuously and fundamentally characterises the ethics of compassion. The paradox is namely that, while proximity, however unspecified, is always required for compassion, the behavior which compassion demands is abstracted from personal conditions. Discussing contemporary problems under the title of an "ethics of broken world reference" (Schulz 1994), he polemicizes them in view of Schopenhauer's ethics of compassion and its applicability.

Compassion considers the other precisely not as a specific person, but as a suffering being as such. The paradox of compassion is that it is directed towards a concrete human being. Yet not on account of his personality but empathizes only because he suffers. Compassion is universal. (Schulz 1994. 71.)

The common ground of utilitarianism and the ethics of compassion also contains the conviction that, if we at all want to understand ourselves as moral beings, we have factually consented to the commandment of eliminating suffering as the elementary moral norm (Tugendhat 1997. 183). How far this consent really goes and how deeply it is anchored in our moral self-conception becomes apparent when this consent is challenged, and particularly when suffering, which could have been prevented, was not prevented, when it has been deliberately brought about, and above all, when it must be responsibly accounted for (Dalferth 2006. 103).

### III. THE MORAL DUTY TO ELIMINATE SUFFERING

From an ethical perspective, the pairing of the experience of suffering with the imperative to alleviate it is not mandatory (Tugendhat 1984; Wolf 1984; Hallich 1998). Suffering is not *per se* a concept with a normative context presenting it as a candidate for the grounding of morals. Like other candidates, it must pass through the filter of an evaluation. It must pass the test of grounding norms in a way that is inter-subjectively valid and capable of general consent, and yet a glance at the history of medicine or psychoanalysis shows how little it is suited for the challenge. It is no accident that the thought that certain forms of experiencing suffering present the challenge of recognising it as something elementarily connected with fragility as a basic constitution of our lives belongs to the central tenets of the ethical self-understanding of entire disciplines and subjects – just think of palliative medicine (Rentsch 1992). And the conviction that a largely unconscious side of the soul may only show itself in suffering, and that this part of the soul cannot be accommodated or employed anywhere else in our life plans: this concept is connected inextricably with the name Sigmund Freud and with psychoanalysis itself. Anyone who perceives in suffering only the occasion to overcome it wants to abolish suffering (Waldenfels 1986). And anyone who – as is so often the case with postmodern esteem for technological possibilities – merely wants to abolish suffering, must be confronted with the question, formulated here with Bernhard Waldteufels, “whether in stoppering the sources of suffering, one does not also stopper the sources of life, whether one does not simply trade in a bit of potential happiness for a bit of security” (ibid.): in other words, whether one does not abolish life along with suffering, as suffering belongs to life in a fundamental sense. Hartmut Kreß, picking up on a basic objection of Rousseau’s,<sup>1</sup> points out emphatically that the technological

<sup>1</sup> “[N]umberless sorrows and afflictions which are felt in all conditions and by which souls are perpetually tormented: these are the fatal proofs that most of our ills are our own work, and that we would have avoided almost all of them by preserving the simple, uniform, and



vanquishment of suffering first of all creates more suffering and second of all opens up a field of problems which are greatly disputed, morally and philosophically (Kreß 1993. 14–15).

It was none other than Sigmund Freud who denounced the immaturity of this model and suggested alternative proposals to the expression “abolishment of suffering” (Küchenhoff 2005). It speaks volumes that the psychoanalysis of Freudian slant wants to ‘remember’, ‘repeat’, ‘work through’ as well as ‘decode’ and ‘translate’ the specific sufferings of people in therapy instead of approaching the illusion that suffering could be ‘overcome’, ‘abolished’ or made not to have happened (Freud 1981).

#### IV. THE IDEA OF THE GOOD LIFE

Even if suffering or the capacity to suffer generally isn’t worth aspiring to, and that which causes suffering ought not to be, the idea of non-suffering is inherent in this judgement, however indistinct it may be. This idea cannot be grasped on the basis of external observation of naturalistic descriptions: such an idea cannot even occur at the descriptive level of a given sense-data which should be summed up externally as one fact among others.<sup>2</sup> It is much more out of the perspective of being impacted and being addressed in which this idea comes into view. The concept of suffering is philosophically relevant to the extent that it refers to possibilities for living which are enclosed within a condition of suffering, but should remain open even in its absence.

It is this openness within the phenomenon of suffering itself whose tension Schopenhauer spells out according to its two sides: illusion-less recognition of the negativistic hardness of that which suffering reveals as a truth about the “existence of the world”<sup>3</sup> on the one hand; but to reach out from this recognition

---

solitary way of life prescribed to us by Nature. If she destined us to be healthy, I almost dare affirm that the state of reflection is a state contrary to Nature and that the man who meditates is a depraved animal.” (Rousseau 1992. 23.)

<sup>2</sup> See the project sponsored by BMBF at the Centre for Interdisciplinary Ethics of the Albert-Ludwigs-Universität Freiburg: *Zu den ethischen Grenzen einer präferenzorientierten Medizin – Eine interdisziplinäre Untersuchung am Beispiel der Anti-Aging-Medizin* under the direction of Giovanni Maio, Lore Hühn, Holger Gothe and Georg Marcmann.

<sup>3</sup> “At base it is redundant to argue whether there is more good or evil in the world: for the mere existence of evil decides the matter, as the good present along with it or after it can never wipe it out and therefore cannot ever compensate for it [...]. That thousands have lived in happiness and bliss: this could never nullify the angst and death pangs of a single individual; just as little as my current wellness can undo my previous suffering. If then there were even a hundred times less ill in the world than is the case, its mere existence would yet be sufficient to ground a truth which may be expressed in different ways, although always somewhat indirectly, namely, that we should not be gladdened, but rather dismayed by the existence of the world; – that the world’s inexistence would be preferable to its existence; – that the world is something which ought not to be; etc.” (Schopenhauer ZA 2. 674–675).

to a horizon of possibilities characterized by its radical capacity to be otherwise on the other. This horizon testifies to its reality within the world in three ways, namely in saying ‘No’ to that-which-is in order to bring resistance into play if possible; but also in the form of a somewhat open place<sup>4</sup> which vouches for the possibility of life as it ought to be or even the good life (Hühn 2002. 163); and thirdly in the form of a fundamental question of what sort of life we factually lead under the dominion of the Will in Schopenhauer’s terminology, and what kind of life we actually want to lead.<sup>5</sup> It is Franz Rosenzweig who viewed this fundamental question as Schopenhauer’s pioneering discovery for the new approach of post-idealistic philosophy. In the effort to explain the essence of that which ought not to be precisely through the phenomenon of suffering, an affinity becomes apparent.<sup>6</sup>

<sup>4</sup> Arthur Schopenhauer, taking up a basic figure from the first print/version of Schelling’s *Ages of the World*, characterized this place as the place of a freedom which lives “above all being,” and further defined this freedom as a freedom from the compulsion to self-realization. The relevant passage from Schelling reads: “The true, the eternal freedom lives only above being. Freedom is the affirmative concept of eternity or that, which is above all time. To most, as they have never sensed this highest freedom, it seems most high to be a being (*Seyendes*) or a subject. They thus ask: what can be thought above being? And they answer themselves: nothing, or something similar to it. Certainly it is a Nullity (*ein Nichts*), but as ‘pure’ freedom is a Nullity (*ein Nichts*); as the Will which wants nothing, which does not desire any matter (*Sache*), for whom all objects (*Dinge*) are equal, and which is therefore not moved by any of them. Such a Will is nothing and everything. It is nothing to the extent that it neither desires to become effective, nor yearns for any form of reality. It is everything because all power (*Kraft*) comes from it alone as the eternal freedom, because it has all things beneath it, commands all and is comanded by none.” (Schelling WA. 26–27.) Schopenhauer comments appropriately: “Rather, we profess it freely: that which remains after the total negation (*Aufhebung*) of the Will is, to all those who are not filled with the Will, certainly nothing. But in reverse, it is for those in whom the Will has taken a turn and negated itself, this our so very real world with all its suns and Milky Ways – nothing.” (Schopenhauer ZA 2. 508.) See on this topic Hühn 2006. 154–155.

<sup>5</sup> “Yet now that we ask about the freedom of the *Will* itself, the question would as such present itself like this: ‘Can you also *will* what you will (*wollen was du willst*)?’ – which emerges as if the Will were dependent upon another Will lying behind it. And if one posits that this question would be affirmed, then a second would also emerge: ‘Can you also will what you want to will (*was du wollen willst*)?’ and so it would be postponed (*hinaufgeschoben*), in that we would always think *one* Will as dependent upon one which was earlier and more deeply situated, and would strive in vain to arrive in this way at a final Will which we would not need to think of as dependent.” (Schopenhauer ZA 6. 46.)

<sup>6</sup> “Schopenhauer was the first of the great thinkers to ask not after the essence, but after the value of the world. A highly unscientific question, if it was really so intended, that one should not ask about the objective value, the value for ‘something’, the ‘sense’ or ‘purpose’ of the world – which would simply be an expression for the question about the essence –, but if the question only referred to the value for humans, or perhaps for the human Arthur Schopenhauer. And this is the way it was intended. The question was consciously referred to the value of the word for humans, but even this question had its fangs broken, in that its solution was once more found in a system of the world. For system already means independent general validity. Thus, the question of the pre-systematic person found its answer in the saint of the closing section, which had been developed through the system. Nonetheless, even this is something unheard of for philosophy: that a type of person and not a concept closed the

It is Arthur Schopenhauer and his disciple Theodore W. Adorno who direct their analysis of that which ought not to be through the phenomenon of suffering and directly raise suffering to the standard of knowledge. With his famous phrase, according to which “the need to give suffering a voice is the condition of all truth” (Adorno 1966. 27), he explains as Schopenhauer before him, that it is precisely the experience of suffering which presents the challenge of morality, at least for one’s own.

The phenomenon of suffering is accorded a key role to the extent that it is inscribed with a normative content from which the ideal of that which isn’t but which ought to be draws a good deal of its evidence. It is the ideal of what out to be which releases the view to the possibility of a good life precisely in letting us say no to that which is suffered. Clearly, this ideal also attests its presence in the decoding of the unconscious appeal which expresses itself in suffering.

#### V. THE COGNITIVE AND INTERPRETATIVE POTENTIAL OF THE CONCEPT OF SUFFERING FOR THE ETHICS OF COMPASSION

No doubt, the cognitive and interpretative potential of the concept of suffering reaches much further than the philosophically relevant way the concept is applied, at least in the narrow confines of Schopenhauer’s ethics of compassion and Adorno’s *Negative Dialectics*. These are certainly forms of suffering which in no way evoke the practical impulse to soothe suffering, nor should they. Examples from medical ethics suffice to demonstrate how necessary it is to separate suffering as an expression of a purely subjective and private experience from its overdrawn normative portrayal, in other words to clarify whether and to what extent suffering can be called upon as something which ought not to be in a morally significant sense (Maio 2006). The fact that it is shameless to deal with suffering in the form of an immediate experience of pain – such suffering as evidenced by a sprained ankle, for instance – in the same breath where one evokes the abyss of nameless suffering in Auschwitz, as Hans Jonas so urgently reports in his *The Concept of God after Auschwitz* (Jonas 1995); all this does not need any further commentary. In short, language presents us with a single concept for a vast and highly complex spectrum of different experiences of pain and suffering in experiences which a concept can scarcely cover without further differentiation. It stands to reason that the complex and many layered phenomenon of suffering can only be joined under a common denominator at the price of illegitimate unities and possibly maddening equivalences.

---

system, really closed the system as its final stone and was not added as an ethical decoration or an appendage.” (Rosenzweig 1988. 8–9.) See also Schwab 2006. 332–334.

Even in a sense which comes close to our quotidian intuition, the concept of suffering stands out on account of its wide range of different meanings and connotations: The unity of these meanings is not apparent without further explanation, and ways of using the concept which may be normatively strong or weak run up against one another unanalysed. In view of the variety of problems which arise here and to avoid making the field which should be analysed fully unfathomable, a rigorous selection is necessary. The question of what suffering is should be discussed in the following with a view to make the discourse about the ethics of suffering assessable. This demands first of all that one brings a certain order to the ambiguity and complexity of the concept of suffering. Naturally, there is no fixed scale for suffering.

As I have shown, the ethics of compassion does not orient itself on just any feeling, but specifically on one in which the subject comports himself to suffering through being impacted by it and thus takes a stance towards suffering that is as follows.

- (1) The stance revealed in compassion relates to suffering first of all in the sense of a negative evaluation as something which ought not to be and certainly ought not to continue in the future. This concept of suffering is not neutral in the sense of putting up with that which one encounters, and cannot be translated as a form of passivity: a passivity, that is, which lives in opposition to activity, although they translate two moments of an opposition which are structurally and logically equal. Such a neutral concept of suffering only plays a role in the ethics of compassion to the extent that it precedes the determination of suffering in the sense of that which ought not be (Angehrn 2006).
- (2) The stance revealed in compassion is secondly one which requires a readiness to alleviate and prevent suffering just as much as it calls again to the impulse to oppose oneself to suffering wherever possible. This concept of suffering is paired with the moral imperative to alleviate suffering, although action wins the upper-hand in the interplay of action and suffering. This leads to a process in which suffering increasingly takes on the character of an evil which should be avoided and evicted in such a way that an end to the negative condition of powerlessness in which the object of compassion finds herself.
- (3) The stance revealed in compassion is thirdly one which conceives of the practical impulse to intervene in a normative way. It does this so deeply that one expects a plausible explanation of moral action as such. Schopenhauer formulates the question of how it is possible to want to alleviate the suffering of others as the question of how the “welfare and grief” (Schopenhauer ZA 6. 247) of others can become a motive for action.

Obviously only in that the other becomes the final end of my Will, just as it is otherwise myself: thus in that I want his welfare with complete immediacy and do not want his grief, I directly suffer with him in his grief, feel his grief as usually I only feel my own, and therefore immediately desire his wellness, as I usually only desire my own. This demands, however, that I am in some way identified with him, i.e. that the entire *difference* between myself and every other, on which my egoism rests, is to a certain degree abolished. (Ibid. 247–248.)

In this definition, compassion is oriented by action and not a state of indifferent, purely aesthetic contemplation. The ethics of compassion appeals to the character of moral obligation which is unique to the concept of suffering in this third definition. For Schopenhauer, this obliging character is placed in parenthesis with a warning – a warning which differs fundamentally from that of Rousseau’s conception to the extent that the latter sanctions a border in compassion as the “natural repugnance to see any sensitive Being perish or suffer” (Rousseau 1992. 15). This is precisely the border which Schopenhauer attempts to eradicate. Schopenhauer undermines the demarcation of the self with reference to the affinity of all beings capable of suffering which comes to light in compassion, while Rousseau does everything he can not to damage this demarcation line. In anticipation of one of Nietzsche’s key figures, Schopenhauer seeks the dissolution of the boundaries of personal relation to the self-evident in the *principium individuationis*, while Rousseau provides it with a stable foundation.<sup>7</sup> Suffering is ultimately essential to life, says Schopenhauer, although this equalisation makes suffering one of the forms in which we carry out our lives. This form extends to the total constitution of the concept of suffering. For compassion should reveal to our eyes the other as a particular form of the same, although the other is only experienced as a particular, transient, individuated form of the Will. Here, compassion becomes a privileged place of knowledge which renders the identity of the Will transparent in all its apparitions. The concept of suffering is thus a border concept which sets the abolition of the difference between ego and non-ego in motion through the reconstruction of the other to my equal.

<sup>7</sup> Though Schopenhauer refrains from naming this difference (see Schopenhauer ZA 6. 285–288), his confrontation with Rousseau shows how unequally compassion is weighted in the balance between compassion and self-love: “[A]nd as long as he does not resist the inner impulse of commiseration, he will never harm another man or even another sensitive being, except in the legitimate case where, his preservation being concerned, he is obliged to give himself preference” (Rousseau 1992. 15).

This occurrence [of compassion], however, is astonishing, even mysterious. It is, in truth, the great mystery of ethics, their originary phenomenon, and the border stone beyond which no metaphysical speculation dare to venture. In this occurrence, we see the partition separating being from being according to the light of nature removed, and the non-ego in a certain way becomes an ego. (Schopenhauer ZA 6. 248.)

The removal which Schopenhauer advises is eloquent to the extent that the equalising of the universalising elements with which the Danziger lends all beings capable of suffering the status of an object of moral consideration comes all too unconditionally to light. He obviously accepts that the concept of individuality, in particular the otherness of the other, is robbed of its entire substance and the other is reduced to an ephemeral apparition of the Will. A further price must be paid: qualitatively different forms of suffering – such as the reflected suffering of human on the one hand or the possibly immediate creaturely sufferings of an animal on the other – may be descriptively differentiated from one another, but are accorded an equivalent normative status. This normative equivalence is connected to the fact that, while there can be different qualities of suffering, these qualities are normatively irrelevant in view of the question of whom we ultimately owe moral consideration (Wolf 2004. 75–85).

Anyone who tries, as is often the case in contemporary debates about animal ethics, to stylize Schopenhauer as the primary witness of an ethics which extends the whole of its universalising foundation along the circle of moral addressees to the animals, must remember that animals attain the position of moral addressee because and only because they are contained in a whole which reaches beyond humans and animals: the Will – the dominant structural principle of reality (ibid. 137–145).

(4) The stance revealed in compassion is fourthly one which accords the affective impact of the suffering of others with the power to conclude truths – a power which refers to the total constitution of human existence and even existence as such. Using the affinity of all beings gifted with senses which shows itself in compassion, Schopenhauer provides a plausible proof that life, in its totality, stands under the sign of the negative – sign that does not come to light anywhere other than the experience of suffering and which marks the way to the metaphysical interpretation of existence according to which the world is fundamentally something which ought not to be.

Yet compassion is not the only experience from which, for Schopenhauer, light falls on the total metaphysical constitution of existing and existence. The impulse announcing itself in compassion is action-oriented and must as such be distinguished from a comportment to suffering which emerges from aesthetic distance. It is, in the original sense of the word, the theoretical stance of observation (Figal 2006) which most clearly opposes the action-oriented impulse in the phenomenon of compassion. And this is the basic contradiction which, and this

is my guiding thesis, shakes Schopenhauer's philosophical consideration about morals to its core. The indifference and composure demanded of the ascetic's contemplative comportment negates that which the ethics of compassion seeks as a whole: namely, to behave in a way which is action-oriented and relates to the suffering of the other in a normatively meaningful way.

Particularly in the tragic form of compassion, the aesthetic subject adapts a certain comportment to the suffering of the protagonists – a comportment which can no longer be appropriately described as an identificatory projection into the mental sensitivities of the other. Rather, through the contemplative comportment of *aesthetic* indifference, it disempowers that which the normative comportment of active compassion, *linked to action*, seeks to attain. 'Tragic compassion' frees the view for such knowledge and has the unveiling of the other as my equal for its goal (Recki 2006). Structurally analogous to the variety of suffering which the tragic hero represents, Schopenhauer says the following of the ascetic who releases himself from the pressure of the Will:

Such a person who, after many bitter battles against his own nature finally overcomes the Will, remains solely as a pure knowing subject, as an unclouded mirror of the world. Nothing can frighten him any longer, nothing move him. [...] Calmly and smilingly, he looks back on the phantasm of this world, which was once capable of moving and tormenting his mind. [...] Life and its forms but float as yet before him, like a transient apparition. (Schopenhauer ZA 2. 483.)

In summary, one may retain the following of the fourth point: that which is revealed in the neutrality towards action of the aesthetic comportment towards suffering becomes the defining factor of an ethical indifferentism which arrests *every* form of action, even those which refer to the "welfare and grief" (ZA 6 247) of the other. Those who attempt, like Schopenhauer, to make the subject's battle for their self-preservation the unalterable reference point for the analysis of the existence of human selves must be confronted with the question of whether such a one-sided determination is at all able to leave room to differentiate between actions which are morally prescribed and those which are reprehensible. It is not by chance that this suspicion has proved itself so hard that a morality founded in this way on compassion puts its own foundation in question without fail (Wischke 1994. 113–114). Finally: how can it meaningfully be maintained that an empirical morality which is anchored in the feeling of compassion is indispensable for ethics, when the whole point of this ethics is to generally obfuscate the empirical setting in which this morality can become effective? And if a moral act, for Schopenhauer, is the negation of self-interest, then such an action is from the start a contradiction which consists in defining itself in opposition to the requirement to which it owes its execution, namely to will the welfare of others.



(5) The stance revealed in compassion is fifthly one which mobilises an immemorial feeling which precedes all reflection in order to preserve suffering's anthropological dignity from dishonest appropriation and to re-establish it. To be sure, it must be protected from the various ways it can be disempowered and cynically belittled through the dominion of a morality which orients itself on reason, but also from any and all attempts of theodicy which would diminish the reality of suffering. Following Rousseau, there is a „repugnance to see any sensitive Being perish or suffer, principally those like ourselves” (Rousseau 1992. 57), which expresses itself in compassion. This is simultaneously the expression of the originary *feeling of relation* which unites humanity with one another along with the animals and which cannot be meaningfully challenged. Far from even noting the asymmetrical difference between the subject and object of compassion, Rousseau, in his thoughts on moral philosophy, takes into account the eminent vulnerability and fragility of all beings capable of suffering. He does this in such a way that, upon immediately observing the sufferings of another, the mind is at once directed to the way one's own nature, as well as that of every other human being, is structured (Ritter 2005. 55–56). Rousseau writes:

There is, besides, another Principle which Hobbes did not notice, and which – having been given to man in order to soften, under certain circumstances, the ferocity of his amour-propre or the desire for self-preservation before the birth of this love – tempers the ardor he has for his own well-being by an innate repugnance to see his fellow suffer. [...] I speak of Compassion, a disposition that is appropriate to beings as weak and subject to as many ills as we are; a virtue all the more universal and useful to man because it precedes in him the use of all reflection; and so Natural that even Beasts sometimes give perceptible signs of it. (Rousseau 1992. 36.)<sup>8</sup>

With his conception of compassion, Rousseau undermines the unbridgeable difference between humans and animals which the Cartesian tradition purposefully stresses in denying animals reason along with the capacity for language, as a result of which they are excluded from the universe of moral addressees. The ethics of compassion has its anti-Cartesian impact in building on a pre-reflexive self-understanding – a self-understanding which is centred in a chosen feeling of relationship and which permits access to that which firstly precedes reflection as much as it receives from it, and which secondly be brought to language in its spontaneity, yet in a way that at the same time is not unfeigned. And yet – this is the discovery for the critique of language – every language which is subordinated to the rules of writing risks blocking the articulation of authentic experience in a fundamental way. Rousseau declares the mute language of gestures and signs to be the depraved form of articulation submerged in a society which sepa-

<sup>8</sup> Translation slightly modified.



rates humans and animals. Such a language belongs to the rules of reason which “reason is later forced to re-establish upon other foundations when, by its successive developments, it has succeeded in stifling Nature” (Rousseau 1992. 15).

#### VI. ROUSSEAU’S ETHICS OF COMPASSION AND THE DIALECTICS OF THE ENLIGHTENMENT

With these lines, Rousseau laid down a sketch of the basic figure of the *Dialectic of Enlightenment*, completed behind the backs of its actors from which the whole history of this figure can be traced. It is the thought of a ‘return of the repressed’ with which Rousseau attempts to illustrate the binding and structuring potency of the state of nature with his “natural feeling” of compassion. This basic thought, which confronts civilised society with the repressed potential of its origins, has its entire point in the fact that it challenges this civilisation so thoroughly that this can ultimately be interpreted as a process of liberation which threatens to collapse into the pathological distortion of the damage that human existence does to itself and through itself. No doubt either that it is the strength of the structural model of the return of the repressed to orient itself explicitly according to the logic of ruinous self-exclusion and as a result to be predestined to offer a framework for explaining and analysing the phenomenon of alienation in all its abysmal negativity. Long before this model established itself through the winding pathways of philosophical tradition of the unconscious to a psychoanalysis of Freudian provenance, thereby advancing itself to one of the key figures in the *Dialectic of Enlightenment* of our days, it served as a guiding concept of cultural criticism which anticipated the critiques of alienation in the 18<sup>th</sup> and 19<sup>th</sup> centuries. Rousseau, as well as Schiller<sup>9</sup> in his footsteps, may have contributed to the career of a model which grew greatly in significance by the end of its diversified career to the extent that the general context of alienation is inescapable for the modern era.

Characteristic for this dialectic is that it aids in the form of a return on an unaccustomed radical level, at least in the modus of extreme alienation in bringing that which was previously repressed to reality, along with all natural feelings, as nature (Horkheimer–Adorno 2004).

<sup>9</sup> In his letters *On the Aesthetic Education of Man*, Friedrich Schiller spelled out the dialectics of the Enlightenment once more through a perspective on the difference between the wild and the barbaric as defined by Rousseau. Schiller writes in his fifth letter: “The enlightenment of the understanding, of which the cultivated classes are not incorrect in boasting, demonstrates all in all so little an ennobling influence on the ethos (*Gesinnung*) that it rather secures its corruption (*Verderbniß*) through maxims. We deny nature on the terrain that justly belongs to it only to experience its tyranny in the moral terrain, and by resisting its impressions, we accept its axioms.” (Schiller 1962. Bd. 20. 320.)

It is very certain, therefore, that compassion is a natural feeling which, moderating in each individual the activity of love of oneself, contributes to the mutual preservation of the entire species. It carries us without reflection to the aid of those whom we see suffer; in the state of Nature, it takes the place of Laws, morals, and virtue, with the advantage that no one is tempted to disobey its gentle voice; it will deter every robust Savage from robbing a weak child or an infirm old man of his hard-won subsistence if he himself hopes to be able to find his own elsewhere. (Rousseau 1992. 37.)<sup>10</sup>

Of course, along with the position on the suffering revealed in compassion, comes the fact that the complexity of the way it can be accessed reacts to the complexity of the concept, of which the concept of compassion so deeply anchored in European philosophy and widely spread in the theory of natural law, particularly Rousseau's, can present but *one* abbreviated expression. The history of this concept includes a spectrum of meanings. On the one hand the Aristotelean theory of affects, so influential through the history of its reception, holds that "compassion is first of all a kind of pain on the basis of something held to be an ill that is pernicious or painful with someone who has not deserved that such a thing come upon him and of which one (secondly) can expect that oneself or one of one's kin may suffer it and this is (thirdly) the case when it seems near" (Aristotle, *Rhetorics* 1385b). On the other hand, one should also mention the tradition of *Misericordia* in the sense of the virtue of Christian charity – a tradition which dominated medieval Christian philosophy. And if, to intensify the argument, the concept of suffering unites an entire palette of heterogeneous meanings, whose unity can only be ascertained with difficulty, under its banner, then the same goes for compassion. It answers the experience of suffering first in the form of a reactive affect which cannot be contained in direct experience, but it is mediated by forms of processing in the horizon of self-interpretation– forms of processing which react to the existential sensitivity and the eminent vulnerability of our lives (Angehrn 2006). Finally, I would like to resume by saying that the concept of suffering is something we approach through interpretation.

## REFERENCES

- Adorno, Theodor W. 1966. *Negative Dialektik*. Frankfurt/M, Suhrkamp.
- Angehrn, Emil 2006. Das Leiden und die Philosophie. In Lore Hühn – Philipp Schwab (eds.) *Die Ethik Arthur Schopenhauers im Ausgang vom Deutschen Idealismus (Fichte/Schelling)*. Würzburg, Ergon-Verlag. 119–131.
- Aristoteles 2002. Rhetorik. In Hellmut Flashar (ed.) *Aristoteles: Werke in deutscher Übersetzung*. Darmstadt, Wissenschaftliche Büchergesellschaft.

<sup>10</sup> Translation slightly modified.

- Birnbacher, Dieter. 2006. Nahmoral und Fernmoral. Ein Dilemma für die Mitleidsethik. In Lore Hühn; in collaboration with Philipp Schwab (ed.) *Die Ethik Arthur Schopenhauers im Ausgang vom Deutschen Idealismus (Fichte/Schelling)*. Würzburg, Ergon-Verlag. 41–58.
- Dalferth, Ingolf U. 2006. *Leiden und Böses. Vom schwierigen Umgang mit Widersinnigem*. Leipzig, Evangelische Verlagsanstalt.
- Das, Veena 1995. Anthropologie des Scherzes. *Deutsche Zeitschrift für Philosophie*. 47. 817–832.
- Figal, Günter 2006. Wollen und Betrachten. In Lore Hühn; in collaboration with Philipp Schwab (eds.) *Die Ethik Arthur Schopenhauers im Ausgang vom Deutschen Idealismus (Fichte/Schelling)*. Würzburg, Ergon-Verlag. 161–169.
- Freud, Sigmund. Erinnern, Wiederholen und Durcharbeiten. In *Gesammelte Werke*. Frankfurt/M, S. Fisher Verlag. Vol. 10. 126–136.
- Jonas, Hans 1995. *Der Gottesbegriff nach Auschwitz. Eine jüdische Stimme*. Frankfurt/M, Suhrkamp.
- Hallich, Oliver. *Mitleid und Moral. Schopenhauers Leidensethik und die moderne Moralphilosophie*. Würzburg, Königshausen und Neumann.
- Hühn, Lore 2002. Die Wahrheit des Nihilismus. Schopenhauers Theorie der Willensverneinung im Lichte der Kritik Friedrich Nietzsches und Theodor W. Adornos. In Günter Figal (ed.) *Interpretationen der Wahrheit*. Tübingen, Attempto Verlag. 143–181.
- Hühn, Lore. 2006. Der Wille, der Nichts will. Zum Paradox negativer Freiheit bei Schelling und Schopenhauer. In Lore Hühn; in collaboration with Philipp Schwab (eds.) *Die Ethik Arthur Schopenhauers im Ausgang vom Deutschen Idealismus (Fichte/Schelling)*. Würzburg, Ergon-Verlag. 149–160.
- Hühn, Lore 2007. Das Mit-Leid. Zur Grundlegung der Moralphilosophie bei J. J. Rousseau und Arthur Schopenhauer. In Nina Gülcher – Irmela von der Lühe (eds.) *Ethik und Ästhetik des Mitleids*. Freiburg, Rombach. 113–133.
- Kreß, Hartmut 1993. Menschliches Leiden zwischen technischer Weltbewältigung und theologischer Deutung. *Pastoraltheologie*. 83. 12–31.
- Küchenhoff, Joachim 2005. *Die Achtung vor dem Anderen. Psychoanalyse und Kulturwissenschaften im Dialog*. Weilerswist, Velbrück Wissenschaft. 345–352.
- Maio, Giovanni 2006. Die Präferenzorientierung der modernen Medizin als ethisches Problem. *Zeitschrift für Medizinische Ethik*. 52. 339–354.
- Nietzsche, Friedrich 1999. Zur Genealogie der Moral. In Giorio Colli – Mazzino Montinari (eds.) *Sämtliche Werke: Kritische Studienausgabe in 15 Bänden*. München, Deutsche Taschenbuch Verlag. Vol. 5.
- Recki, Birgit 2006. Mitleid ohne Freiheit? Überlegungen zu Schopenhauer und Kant. In Lore Hühn (in collaboration with Philipp Schwab) (eds.) *Die Ethik Arthur Schopenhauers im Ausgang vom Deutschen Idealismus (Fichte/Schelling)*. Würzburg, Ergon-Verlag. 23–29.
- Rentsch, Thomas 1991. Philosophische Anthropologie und Ethik der späten Lebenszeit. In Paul B. Baltes – Jürgen Mittelstraß (eds.) *Zukunft des Alterns und gesellschaftliche Entwicklung*. Berlin, De Gruyter. 283–304.
- Ritter, Henning 2005. *Nahes und fernes Unglück. Versuch über das Mitleid*. München, Beck.
- Rosenzweig, Franz 1988. *Der Stern der Erlösung*. With an introduction by Reinhold Mayer and a *Gedenkrede* by Gershom Scholem. Frankfurt/M, Suhrkamp.
- Rousseau, Jean Jacques 1992. *Discourse on the Origins of Inequality (Second Discourse). Polemics and Political Economy*. Translated by Judith R. Bush – Roger D. Masters – Christopher Kelly – Terence Marshall. Hannover, Dartmouth University Press.
- Schelling, Friedrich W. J. 1946. *Die Weltalter Fragmente [= WA]*. Published by Manfred Schröter. Munich, Leibniz Verlag.
- Schiller, Friedrich 1962. Über die ästhetische Erziehung des Menschen in einer Reihe von Briefen. In Benno Wiese et al. (eds.) *Schillers werke. Nationalausgabe*. Weimar, Böhlau. Vol. 20. 309–412.

- Schopenhauer, Arthur 1997. *Werke* [=ZA]. With the collaboration of Arthur Hübscher, published by Angelika Hübscher. Zürich, Diogenes.
- Schulz, Walter 1994a. *Der gebrochene Weltbezug. Aufsätze zur Geschichte der Philosophie und zur Analyse der Gegenwart*. Stuttgart, Neske.
- Schulz, Walter 1994b. Aporien und Antinomien der Subjektivität und der Versuch, sie „festzustellen“ und „aufzuheben“. Anmerkungen zur Geschichte der philosophischen Metaphysik. In *Der gebrochene Weltbezug. Aufsätze zur Geschichte der Philosophie und zur Analyse der Gegenwart*. Stuttgart, Klett-Cotta.
- Schwab, Philipp 2006. Der Asket im System. Kierkegaards Kritik an der Kontemplation als Fundament der Ethik Schopenhauers. In Lore Hühn (in collaboration with Philipp Schwab) (eds.) *Die Ethik Arthur Schopenhauers im Ausgang vom Deutschen Idealismus (Fichte/Schelling)*. Würzburg, Ergon Verlag, 321–345.
- Spaemann, Robert 2001 Tierschutz und Menschenwürde. In Robert Spaemann (ed.) *Grenzen. Zur ethischen Dimensionen des Handelns*. Stuttgart, Klett-Cotta. 467–476.
- Theunissen, Michael 1983. Negativität bei Adorno. In Ludwig von Friedenburg – Jürgen Habermas (eds.) *Adorno-Konferenz*. Frankfurt/M. 41–65.
- Tugendhat, Ernst 1984. Drei Vorlesungen über Probleme der Ethik. In *Probleme der Ethik*. Stuttgart, Reclam. 57–131.
- Tugendhat, Ernst 1997. *Vorlesungen über Ethik*. Frankfurt/M, Suhrkamp. 177–196.
- Waldenfels, Bernhard 1986. Das überwältigte Leiden. Eine pathologische Betrachtung. In Willi Oelmüller (ed.) *Leiden* (= Kolloquium zur Gegenwartsphilosophie, Bd. 9). Paderborn, Schöningh. 129–140.
- Waldenfels, Bernhard 2000. *Das leibliche Selbst. Vorlesungen zur Phänomenologie des Leibes*. Frankfurt/M, Suhrkamp.
- Wischke, Mirko 1994. *Die Geburt der Ethik. Schopenhauer – Nietzsche – Adorno*. Berlin, Akademie Verlag.
- Wolf, Ursula 1984. *Das Problem des moralischen Sollens*. Berlin – New York, De Gruyter. 148–199.
- Wolf, Ursula 2004. *Das Tier in der Moral*. Frankfurt/M, Klostermann.

HYE YOUNG KIM

## An Existentialist Analysis of Forgiveness and Gratitude

Gratitude and forgiveness are not the most frequently summoned topics in the context of existentialist emotions. One can find some literature on forgiveness and gratitude mostly in ethics, theological philosophy, or psychology, but not much from the existentialist perspective. Why? There are a couple of problems when dealing with these in the context of existential understanding. First, there is a question that relates to the definition and the categories of emotions, in other words, whether we can deal with gratitude and forgiveness under the auspices of emotions at all. Another question is whether the fundamental understanding of existence can be discussed in interpersonal dimensions.

To investigate these problems in detail, we should first clarify what we understand by existentialist philosophy. Briefly, it is a philosophy that tries to understand human existence. Therefore, the fact that there has not been much literature on forgiveness and gratitude in the existentialist context means that not much has been said about forgiveness and gratitude in the context of understanding our existence. Why is that?

We can infer one of the reasons from the reason why other emotions such as angst, anxiety, guilt, or fear (regardless of the question of whether they are emotions or not) frequently appear in the existentialist discourse. The core of human existence lies in the inevitable fate of human beings – death. It is not only a conventional truth, but it is also *the* existential truth: we all die at one point. One of the ways that we cope with this fate is to try to understand it. Reflecting on the end of existence is accompanied with different emotions from that of reflecting on logical consequences of complicated equations. For the most part, and for most of us, reflecting on death does not usually come with warm feelings such as comfort or joy. Thinking about death, either mine or that of others, usually arouses uncanny, un-homey, anxious, scary, uneasy, or sad feelings. In fact, extremely complicated equations might also cause you anxiety, fear, sadness, or all of these at the same time, in that you fear that you might not be able to solve them, or unsolvable equations might even make you feel like you are going to die, but the difference is that you can give up on the complicated equations or

solve them eventually and set yourself free, but you cannot give up nor solve your death. There is nothing you can do about it. This feeling of helplessness builds the roots of anxiety. In this context, the relation between existential understanding and the uncanny emotions is inevitable.

Another reason that forgiveness and gratitude do not appear in the context of existentialist emotions is because forgiveness and gratitude are interpersonal, while those uncanny emotions, along with the feeling of nothingness are rather grounded in a self-centric, isolated state of being. Emotions such as anxiety, fear, and guilt are felt by a solitary being in the face of their own nothingness. In this context, these emotions were dealt with as non-object-oriented feelings. Therefore, they were differentiated and renamed, for instance, as mood, because they are not involved with other people or things. Death is always only a possibility of my own not-being. The object of these feelings is, therefore, not-being, in other words, it *is* not there. The object is not there, and what is present is only the feelings themselves and my understanding.

But are they emotions? Or feelings? In this text, I have already referred to them as feelings or emotions. I did so because we become conscious of them in the way that they are felt. However, the question of whether they are emotions, or feelings, or both is not only not simple but also controversial. Scholars in the study of emotions do not agree on a unitary definition and categories of emotions and feelings due to their different methods and approaches. However, I am going to use these terms more freely in this essay, despite the fact that they are often strictly distinguished and applied to different situations. This doesn't mean, though, that I regard forgiveness and gratitude as merely feelings or emotions. They are more than emotions or feelings, which I will discuss further, but they are *felt*, or in other words, we feel them when we have them. They affect us *emotionally* and cause us to *feel* something or to situate ourselves in a certain *mood*.

I mentioned that the common character of forgiveness and gratitude is that they happen in an interpersonal context. Namely, it occurs between more than one person. In the ordinary sense, they are object-oriented-emotions. In that sense, gratitude and forgiveness could be considered as secondary emotions in terms of existentialist emotions compared to angst, anxiety, guilt, etc. which I would like to refer to as solitary existentialist emotions. The reason forgiveness and gratitude seem secondary in existential understanding is that the presence of the other is the necessary condition in the case of gratitude and forgiveness. For example, the way Sartre deals with gratitude in the context of "Concrete Relations with Others" in *Being and Nothingness* (1984) is a fragmentary mentioning with other feelings in his analysis of love and hate. I question, though, whether the solitary existentialist emotions are more primary than the interpersonal emotions in existential understanding. There are many different types of interpersonal emotions. Then why forgiveness and

gratitude? In other words, what makes them more significant than other interpersonal emotions in the context of existential understanding? Let's start with forgiveness.

## I. FORGIVENESS

What does it mean to forgive? Or to be forgiven? First of all, grammatically it is a verb. It is an action. And it is a transitive verb which means that it takes an object. One forgives someone; this someone is an object of the action of forgiving. Forgiveness itself is not a feeling, but it causes the one who forgives and the one who is forgiven to *feel* to forgive or to be forgiven, i.e. to be in the state of mind to be able to forgive or to be forgiven. And this feeling is crucial for this action.

We all know what it means to forgive, but in fact we do not really understand what it means. In general, to forgive is understood as a synonym of pardon. It is defined as "to cease to feel resentment against an offender" as in "forgive the enemies," "to give up resentment" as in "forgive an insult," or "to grant relief from a debt."<sup>1</sup> To forgive does not mean to forget, even though we confuse them often in daily life. To forgive means to give up resentment and make room to free the offender and eventually myself, in other words, it is to treat the offender as not guilty. Here lies the core character of forgiveness as an existentialist emotion: making room.

Existentialist philosophy is an attempt to understand human existence. The emphasis is on 'to understand.' In effect, Heidegger's analysis of 'understanding' as human existence (see Heidegger 2006) reveals the quintessence of existentialist philosophy. This understanding as existential understanding (*existenziales Verstehen*) is, however, differentiated from an intellectual understanding, it is rather human existence itself, which means that we exist in the way we question and somehow understand our own being. Therefore, simply being there without existential understanding is not yet existence.

The task is clear: to understand existence. What do we understand by existence? If existence is a *Geschehen* (occurrence), which *geschieht* (occurs) between the beginning and the end (see Kim 2015), the end necessarily holds a special status in the structure of understanding, because a *Geschehen* as a whole can be understood only after it is ended, hence the end is the key to understand this *Geschehen*. In the case of human existence, however, the end is never there yet, because the subject of understanding has to be there to understand. This is a paradox of self-representation. The paradox arises from the situation in which the object of understanding is the existence of the subject of understanding.

<sup>1</sup> Merriam-Webster: <https://www.merriam-webster.com/dictionary/forgive>



One of the ways to solve this paradox is to grasp this end before the actual end, which is possible through fore-grasping.

The *Vorlaufen* (fore-running) to a possible end is indeed a brilliant installation for resolving this paradox to gain the possibility of understanding (see Heidegger 2006). It is not a coincidence that the structure of *Vorlaufen* as *Sein-zum-Tode* (being to death) is situated (*sich befindet*) in angst, because the possible end is the possible end of my being. Envisioning my forthcoming death, the realization of which is induced from the death of others, does not exactly come with soothing feelings of comfort. In effect, that is why we usually don't – and probably shouldn't – think about the possibility of our own death ceaselessly day and night. Usually we live a life “normally” in the midst of the others, or as in the style of fundamental ontology, “*uneigentlich*” (inauthentically) (see Heidegger 2006). Either way, we live in the world thinking not only about our own existence. This is the point where existential guilt enters to achieve an *eigentliches Verstehen* (authentic understanding). Being guilty is the way of awakening the self to face existence as understanding of existence by fore-grasping the end. Now, the story can be understood with its end fore-grasped (see Kim 2015). However, where does this story begin?

The beginning of existence – there are actually two beginnings. One is ontical (*ontisch*) and the other is existential, and there is a temporal-existential difference between the ontical and the existential beginnings. The ontical beginning is the moment of birth, and the existential beginning is each moment, i.e. the present. Strictly speaking, if existence is existential understanding, the beginning of existence is now, the present where the understanding begins. However, the existential beginning is necessarily based on the ontical beginning, in the sense that the ontical beginning is the necessary condition for existence. Simply, at one point, we all have to be born to start existing. The existential is present, yet the ontical beginning has to be restored. The ontical beginning is the past that has to be re-grasped.

When and how did Sisyphus start to roll up the stone? Who or what placed him there? It is clear that one does not even have to wait for the end of this anxiously. This is the point where suicide becomes a serious philosophical question, possibly the only one (Camus 1984). Fore-grasping is not the only option. One can actually grasp the end of existence and end the cycle of absurdity. Ending the existential understanding by grasping the end – not only fore-grasping it – is an equally available option as fore-grasping it. However, the beginning is different. It cannot be grasped. It can only be re-grasped. It cannot be given up, because it has already happened – I am here.

Either accepting it by understanding it, or avoiding it by ending it, we somewhat know how to deal with helplessness in the face of death. But what about the beginning? This is where the ultimate helplessness stems from. I was delivered to the world regardless of my will, desire, awareness, or understanding. I



had no choice. This is another inevitable condition of being. I can either choose to be guilty to face the question of my being, in other words, to be *verantwortlich* for my being (see Heidegger 2006), or not. But, how can I deal with the beginning?

If it cannot be given up, it can either be forgotten or understood. Forgetting is an option, as it is for the end – we can, and actually do, live in the world as “jeder ist der Andere und Keiner ist er selbst” (Heidegger 2006, 128). There should be more though. Not only *either* oblivion *or* anxiety. What do we have for this another *or*?

This vulnerability and ignorance over the beginning is the ceaseless source of fragility of being. However, this beginning is not a one-time event. It lasts. To be thrown in the world is not a simple instant that one passes by once. We are not thrown in the world as a ball is thrown, as the ball that I throw in the air that leaves my hands immediately. We all are attached to and completely dependent on the others, the world, inevitably, *at least* for years. The completely vulnerable state of which I am not even fully aware is the beginning of my existence, which composes a large portion of my life. The reason this complete dependency on the world becomes problematic is because the world, that I cannot help but depend on completely, is so fragile itself, because it consists of each one of us: anxious, finite beings. Therefore, life is often unbearably hard, and at one point you realize that you are thrown here without having had any choice. Why am I here? “Why is there something rather than nothing, including myself?” – is the question.

In fact, we don’t even have to run towards to the possibility of death in order to face the ungraspable nothingness that only *self-nothing-izes* (*sich nichtigt*) and everything slips out of (see Heidegger 2007) to choose to dwell in the fundamental homeless state in angst. We are already placed in the endless void of vulnerability. After all, helplessness is how we all started. Now the question is: what can be done in the midst of this void? From this beginning, we need to leap to the existential beginning. The first thing to do to make this leap is probably not to run towards the end. What needs to be done is to have my place somehow in this void – to have *my now*. We need to create space. The space in which I can start understanding myself to realize that I am there. It is to *give room* to myself to dwell in and *see* my place. This is what I understand by “to forgive my own being.” Forgiving oneself means creating space of understanding by giving oneself room. It is a process of detaching from the world, and at the same time, finding my place in the world.

When I forgive myself for being there-here, I can treat myself as not guilty. This is to give room to land my feet on while standing out (ἔκ-στασις) in the middle of nothingness. The act of self-forgiving is still interpersonal in the sense that I treat myself as another person who is forgiven and separate: I as the for-giver. Self-observation is the beginning of awareness. This is a room-making

process. And self-forgiving is a process of giving this room for self-observation. This is, in fact, not the beginning of existential understanding, but rather is it the end for which the existential understanding aims. If angst is the state of being, forgiveness is *staring at* the being. The look at my othered self can give myself room to start existing. The way one can be responsible (*verantwortlich*) for one's own being is to be guilty by understanding my finiteness (death) and forgiving myself by understanding my vulnerability (birth).

There are, however, two paradoxes of forgiveness. First, forgiving myself is in fact forgiving the unforgivable. Second, forgiving myself for being there is also forgiving others. Let's have a look at the first paradox. Derrida discusses forgiveness in the socio-political context (see Derrida 2001) rather than in existential sense, but his statement on forgiveness, forgiving the unforgivable, penetrates the essence of forgiveness as a crucial moment of existential understanding. Forgiving my existence is forgiving the unforgivable, because it is, in fact, impossible to forgive my existence. It is impossible not because it is an unforgivably immoral sin or the worst crime. When it is said that the necessary condition of existential understanding is to be guilty of my own being, it is not like being guilty of trespassing. It is to forgive the impossible, because *it* is not an object, and neither a person, a thing, nor an action that can be forgiven. There is only an action of forgiveness – forgiving the forgiver itself. This is the first aporia of existential forgiveness.

The second paradox is that existential forgiveness is self-forgiveness, yet interpersonal. It is interpersonal not only between self and othered self but between self and others, because the vulnerability of the ontical beginning is based on my complete dependency on the others. Forgiving myself for being there means forgiving myself for being in the world, where I am with the others. Forgiving myself requires forgiving the others who share the responsibility for my being as well as for theirs. The anxiety of existing is chained with the anxieties of the others. Being with the other in the world makes me not only to be fallen, lost in the they, and forget about my own being, but often they are the ones who make me be aware of my being. If the others are a phenomenon, as Sartre states (see Sartre 1984), this phenomenon is profoundly related to my existence and reflects my existential understanding. Their eye is already – at least somewhat – implanted in my self-observation. Their influence is so overwhelmingly enormous that it takes much effort to separate and understand myself as my own. Forgiving is giving myself room between me and my othered self, and the others, and realize this space of mine. If being guilty existentially means calling myself to awaken my consciousness to observe my own being, forgiving is yielding room for this awareness.

## II. GRATITUDE

So, forgiving is about forgiving myself, giving myself room to seize (or, to understand) my beginning. What's next? You give, then, you receive. That is gratitude.

Normally, being grateful means to appreciate what is given. We think of tears usually as physiological reaction to sadness or frustration. But we observe that people shed tears when they experience unconditional, or sometimes unexpected, generosity or benevolence from others, and feel deeply thankful. Strawson's (1974) account to regard gratitude as a reactive attitude with an essential connection to the practice of holding others to normative expectations (see Manela 2015) points out one crucial aspect of gratitude. The tears shed are a representation of the reaction to the others for their acts, physically realized. The other's presence and the interaction with them lies in the core of the ordinary understanding of gratitude. However, usually these tears come out before my ethical evaluation, moral judgement, or determination to express my feelings of thankfulness. This reactive attitude is rather a reflex response. Gratitude is not always what comes after my evaluation of the situation in accordance with my expectations from the others and my comprehension of the benefits.

In philosophy, gratitude is conceptualized also as emotion, virtue, relationship enhancement, or obligation (see Manela 2015). For example, gratitude has appeared in political philosophy as in the theory of political obligation, such as when Socrates claims his obligation to obey the laws of Athens is an obligation of gratitude for benefits received. In the philosophy of religion, gratitude is related to the moral argument for the existence of God. And there have been debates as to whether gratitude is compatible with a belief in determinism (Walker 1980). In general, gratitude features in a social, interpersonal context that functions to establish and strengthen the mutual bonds between individuals and the community (see Walker 1980). In this respect, however, it is still hard to spot a special place for gratitude in existential understanding.

Existential understanding is inevitably solitary in an idealistic way, for the others are the other subjects outside of my self-understanding. However, the fundamental vulnerability of my existence is essentially related to others, whether I am conscious of them fully as other subjects or as phenomena, as objects of my subject. Somehow, we have to accept this, that they are there as I am there, before running towards the future and fore-grasping my possibility of being. The separation process of myself from the world can, or should, come after my solid understanding of my now in the midst of the world with the others. I say 'after' and 'before,' but they do not indicate their temporal order; they happen simultaneously.

To be guilty is to understand where my now is headed, and to forgive is to understand where my now comes from. And to understand where my now is

between where it came from and where it goes toward, we need to understand how my now is there in the world. My now in the world is there always with the now's of the others.

Gratitude means being aware of this present state, i.e. being aware of being given. I gave myself room to observe my beginning so that I can pre-grasp my end. The moment of here now is given. If forgiveness was to create room, gratitude is to create meaning in the meaningless being-thrown. Gratitude is to gain *Sinn von meinem Sein* out of *Sinn von Sein* – grasping my present, not floating in the angst of nothingness. In this sense, gratitude is more than feeling grateful. It's my relation to the world as a being-in-the-world.

Gratitude, as *gratia* means grace, favor, goodwill, kindness, friendship and *gratus*: dear, beloved, pleasing, acceptable, agreeable. In this regard, gratitude relates to the state of being approved or held in regard. It is a recognition and agreement of the act that was done. To be grateful is to accept and agree with what is given. But it is not a passive moment: it is to approve my present.

Gratitude is interpersonal, yet the core of gratitude lies in self-acceptance. Here gratitude does not mean to be grateful for a specific thing or a person. It is to appreciate my being itself – now. Now is always there but never there. It was the past and will be the future. It is only an *Augenblick*, a constant flip between the past and the future, between constantly flipping moments. It is not really a stable moment. It is not there. But it is there as now when it is within in the network of now's of the others and things, all that there are in the world. It is spatial. Now has no temporal duration, but it has a spatial duration between my now and the now of the others. Gratitude is to become conscious of this now in the world – the awakening. There are multiple complex ways to relate to the others in the world. But gratitude, in this sense, is the primary way of connecting myself to the world and to see and be aware of this connection.

Forgiving and being grateful for my being means accepting existence. Not only understanding it but accepting it. It's not the victory over death, nor the denial of the vulnerability of being, but acceptance of existence as it is. It is a gaze at existence standing here and now, instead of running forwards and backwards trying to understand.

## REFERENCES

- Camus, Albert 1984. *The Mythos of Sisyphus*. New York, Penguin Books.
- Derrida, Jacques 2001. *On Cosmopolitanism and Forgiveness*. New York, Routledge.
- Heidegger, Martin 2006. *Sein und Zeit*. Tübingen, Max Niemeyer Verlag.
- Heidegger, Martin 2007. *Was ist Metaphysik*. Frankfurt/M, Verlag Vittorio Klostermann.
- Kim, Hye Young 2015. *Sorge und Geschichte: Phänomenologische Untersuchung im Anschluss an Heidegger*. Berlin, Duncker und Humblot.
- Manela, Tony 2015. Gratitude. *Stanford Encyclopedia of Philosophy*.  
<URL: <https://plato.stanford.edu/entries/gratitude/#GraEmo>, last access 02.07.2020. >  
Last access 31th March, 2020. >
- Sartre, Jean-Paul 1984. *Being and Nothingness*. New York, Washington Square Press.
- Strawson, Peter Frederick 1974. Freedom and Resentment. *Freedom and Resentment, and Other Essays*. London, Buttler and Tanner. 1–25.
- Walker, A. David M. 1980–1981. Gratefulness and Gratitude. *Proceedings of the Aristotelian Society. New Series*. Vol 81. 39–55.

JAMES CARTLIDGE

## Heidegger's Philosophical Anthropology of Moods

This paper concerns the account of Dasein's existence that Heidegger gives us in *Being and Time*, and exactly what kind of account it is. I will argue that, despite his emphatic insistence to the contrary, it should be read as a philosophical anthropology because it gives an account of human existence and its structures. Heidegger's analytic of Dasein is at its most productive and interesting when understood this way, and the reasons he gives for its being essentially different from philosophical anthropology are unconvincing. Heidegger took great pains to distance his work from philosophical anthropology, repeatedly claiming in numerous texts throughout his career that to understand it as such is a mistake. It is not a mistake: his 'analytic of Dasein' has great potential to benefit the philosophical-anthropological project and constitutes a powerful attempt to describe human existence and account for how it is structured. This can be evidenced in many ways, given the breadth and depth of *Being and Time*, but here I will focus on its analysis of moods.

To begin with, however, I will have to discuss anthropology and explain how philosophical anthropology differs from it. I claim that anthropology's general concern is with giving accounts of specific human societies and understanding human differences. A noble project to be sure, but not the project of the philosophical anthropologist. Where anthropology is preoccupied with specificity and difference, philosophical anthropology is concerned with commonality, with what is universal, necessary and constitutive for human existence in general. Anthropologists pursue questions about, for example, what Balinese people and Balinese society are like,<sup>1</sup> or what primitive societies are like and what concepts should be used to describe them.<sup>2</sup> Philosophical anthropology takes place at a more abstract level, pondering *what being a human is like*, and what is involved in

<sup>1</sup> As Clifford Geertz famously does, in *The Interpretation of Cultures* (1973).

<sup>2</sup> There are numerous examples of such research. Some examples would be Bronislaw Malinowski's *Argonauts of the Western Pacific* (1922) and Pierre Clastres' *Society Against the State* (1972). An example of the more conceptual reflection can be found in Edward Dozier, *The Concepts of "Primitive" and "Native" in Anthropology* (1955).

living a human existence. This discussion will lead me onto Heidegger's conception of his project, and how he frames it as a critique of and distancing from both anthropology and philosophical anthropology.

Heidegger claims he is giving an account of *Dasein* and that, because his project is motivated by more fundamental ideas and concerned with a different topic, his project cannot be philosophical anthropology. It is rather a "fundamental ontology," concerned with *Being in general*, not with any particular type or group of particular beings, like "human beings". Philosophical anthropology, in his terms, would amount to a "regional ontology," accounting for this specific region of beings, not Being in general.<sup>3</sup> But the method Heidegger uses for his investigation raises serious questions about whether this is really the case. As we will see, central to Heidegger's proposed method for answering the question of the meaning of Being is the giving of an account of the existence of the only entity that could ask, understand and answer this question. If we want to answer the question of Being, some kind of account is needed of the questioner's kind of existence and how this existence gives rise to its ability to ask, understand and answer questions. Only then could we really know what answering this question would consist in, since it is only through and out of this kind of existence that the question could even potentially be answered. The nature and structure of this existence, therefore, is something should be thoroughly clarified beforehand.

The entity in question, Heidegger famously calls *Dasein* – not "the human being". This is because even though human beings are *Dasein*, being *Dasein* is not necessarily limited to human beings – there may be other entities that could understand, ask and answer the question of Being, whose existence may be bound by similar structures to ours. According to Heidegger, when he analyses *Dasein*'s existence, this means he is engaged in something more fundamental than and essentially different to philosophical anthropology, because he is precisely not giving an account of human existence, but one of *Dasein*. Crucially, though, *anything that truly applies to Dasein applies truly to human beings, because human beings are Dasein*. To give an account of *Dasein*, therefore, just is to give one of human existence, and so at the very least involves and produces a kind of philosophical anthropology.

Examining Heidegger's account of *Dasein*'s moods is a potential way (of which there are many) that we might see this to be the case. Heidegger gives an interesting analysis of moods that perhaps improves on some previous treatments of the affective dimension of human beings, and perhaps manages to state true things about this dimension. Heidegger casts moods as "fundamentally disclosive", as being integrally involved with our making sense of the world

<sup>3</sup> Heidegger capitalizes 'Being' when he is referring to being in general. Since this may be of help in understanding some passages, I will sometimes do the same.

and our disclosure of the meaning of things in it. They don't just reveal 'how we are doing', they play a part in the disclosure of various important aspects of our existence. They are part of the process by which we disclose of objects as meaningful, it is through them that insights about our being as a whole can be disclosed to us, and they are also revealing of an aspect of our condition that Heidegger calls "thrownness". Now, we do not have to agree with everything Heidegger says about moods, and my sketch of his account of them will be limited. But we do not have to examine what he says in too much detail to realise that his work constitutes an account of human existence, even if this is not all that it does. If Heidegger managed to say anything true of Dasein, he managed to say something true of humans. His account is therefore (at least partially) a philosophical anthropology, and is especially productive when read as one.

#### I. ANTHROPOLOGY AND PHILOSOPHICAL ANTHROPOLOGY

Before we can get to why Heidegger's work should be understood as a philosophical anthropology, we have to understand what this is and how it differs from "regular" anthropology. In a 1929 text, Heidegger gives what I think is a fair definition of anthropology:

Anthropology means the science of man. It embraces all that is knowable relative to the nature of man. [...] Within the domain of anthropology [...] falls not only man's human qualities which, because they are at hand, are discernible and distinguish this determinate species from animals and plants, but also his latent abilities, the differences according to character, race and sex. (Heidegger 1973. 146.)

Anthropology is the academic discipline which studies human beings from various perspectives and aims for scientific rigour in doing so. Because human beings have many different aspects, they offer many different phenomena worthy of study, and anthropology not only studies how human beings are distinct from animals, but also how they are distinct from one another along biological, societal, racial and sexual lines. Anthropology scientifically studies human beings from various perspectives which adopt different practises with varying scopes of enquiry. Over time, the discipline has broken down into (roughly) four main categories: cultural, social, linguistic and biological. I will give a brief sketch of these now, and it will necessarily be limited. However, it will be enough to notice anthropology's salient focusses and how the discipline differs from philosophical anthropology.

Cultural anthropology analyses particular cultures, usually on their own terms and without necessarily comparing them to others. A cultural anthropologist might examine historical evidence in order to compile a theory about, for ex-



ample, what ancient Egyptian culture was like, or the culture of Renaissance France. Social anthropologists attempt something similar, but tend to avoid the term 'culture' because the object of their study is better captured as the analysis and comparison of *social relations* around the world. Where cultural anthropologists may be interested in a particular culture in general, social anthropology focusses on the social relations in a given society – so while the topics that might interest social and cultural anthropologists can be very similar, the approach they take is slightly different, as are the conceptual lenses they use. Linguistic anthropology analyses and compares human language and catalogues information about them across the world and throughout history. Biological anthropology analyses the biological basis of human beings, either in terms of its evolutionary history or its modern manifestations.<sup>4</sup>

With this admittedly cursory glance at anthropology in hand, I think we can notice something about its salient focus – in its analysis of human beings, anthropology is concerned with *specificity* and *difference*. Anthropological studies are almost always concerned with giving accounts of specific, particular societies at specific points in history, or with comparing societies and their structures. Be it through historical analyses, biological investigation or ethnographic research which describes first-hand experiences, anthropologists attempt to understand human societies in their historical specificity, and has amassed a wealth of information to this end. As we understand more and more about what particular human societies are like, we naturally understand more about how they compare to one another and what the differences are between them. The preoccupation with human differences has not gone unnoticed by certain anthropologists. Conrad Phillip Kottak, for instance, named one of his books *Anthropology: The Exploration of Human Diversity*. (Kottak 1997) Ruth Benedict, in her study on Japanese culture, wrote that the “tough-minded” anthropologist’s “goal is a world made safe for human differences” (Benedict 2005. 15). Clyde Kluckhohn wrote similarly: “anthropology provides a scientific basis for dealing with the crucial dilemma of the world today: how can peoples of different appearances, mutually unintelligible languages and dissimilar ways of life get along peaceably together?” (Kluckhohn 1949. 1). By understanding what specific societies are like, we can also get an understanding of how these societies and peoples differ from one another, and understanding human differences is crucial to being able to live peacefully in spite of them.

It is on this point of specificity and differences that we can delineate the project of philosophical anthropology, which I understand to be the other side of anthropology's coin. Rather than focussing on specific societies and human

<sup>4</sup> I have not mentioned archaeology here, although it is sometimes said to be a kind of anthropology. Certainly it is involved in the investigation of human cultures and often provides historical evidence for anthropologists to use in their investigations.

differences, philosophical anthropology attempts to find commonality in all instances of the human experience, what is universal, necessary and constitutive for human existence regardless of which society a person lives in. Naturally, this inquiry can take many forms. Philosophical anthropologists aim to specify and elucidate the structures of human existence, which often takes the form of searching for essential and unique features of human beings, or non-essential and non-unique ones that are just particularly important. Their analysis, therefore, could potentially take place from almost any perspective within philosophy: metaphysics, ethics, epistemology, political philosophy, and so on – all of these areas of inquiry could involve or lead to the giving of some kind of account of human existence, and so could be philosophical-anthropological in nature, even if this is not what they are concerned with doing all the time. Whatever method taken, or perspective inquired from, philosophical anthropology gives an account of the universal, necessary and constitutive structures of human existence, considering what it is like to live a human life. Heidegger therefore defines it also quite adequately when he says that philosophical anthropology is “an essential consideration of the human being [...] thereby to work out the specific, essential composition of this determinate region of beings. Philosophical anthropology therefore becomes a regional ontology of human beings” (Heidegger 1973. 148). However, for technical reasons related to his own project, this is exactly what he wants to distinguish himself from, and it is in this notion of a “regional ontology” and his delineation of his own project as “fundamental ontology” that we can see why.

## II. WHAT IS FUNDAMENTAL ONTOLOGY?

Heidegger argued that his project is one of “fundamental ontology”, which makes it essentially different to philosophical anthropology because fundamental ontology is singularly concerned with *the question of the meaning of Being*. This question was the heart and soul of Heidegger’s entire career and he was convinced that his project, being motivated and oriented in this way, meant that he could not be doing philosophical anthropology – he was, by his own estimation, engaged in something different and far more fundamental. He had no time for those who misunderstood this, especially in his infamous private “black notebooks” where he wrote, for example, that “if the question of being had been grasped, even if only in a crude way [...] then *Being and Time* could not have been misinterpreted and misused as an anthropology” (Heidegger 2016. 16). He goes even further elsewhere: “anthropology is the preventive measure instituted by modern humanity in consequence of which the human being arrives at not wanting to know who he is” (Heidegger 2017a. 18). Heidegger even calls a writer who was influenced by his work (Otto Bollnow) a “philistine” for

“tak[ing] it as settled that *Being and Time* is a philosophical anthropology” (Heidegger 2017b. 170). These are just a selection of many examples, public and private, of Heidegger making such criticisms: it was a persistent problem for him that people “misunderstood” his work in this way, and one he could not have been clearer about wanting to repudiate.

Perhaps in the context of his career, and the project of *Being and Time*, Heidegger's exasperation is somehow understandable. After all, it goes to the very heart of his work and the kind of sweeping criticisms he constantly made about basically every significant philosopher of the Western tradition except himself. Those familiar with Heidegger will no doubt have examples of this in mind: he often, repeatedly claimed that philosophers, from Plato to Aristotle, Descartes to Kant, Hegel and beyond, had all been somehow mistaken, however valiant their efforts. Why? Because they failed to adequately address the question of the meaning of Being – the most important question of them all. Heidegger often referred to it as the “grounding-question”, such as here, where he is also talking about misinterpreting his work as anthropology: “this misinterpretation is basically excluded [...] if from the beginning we hold on to the grounding-question of the meaning of being as the only question” (Heidegger 1999. 60).

But why is this question so important? Being is fundamental to everything we do, every sentence that we speak, everything that *is*, but “we do not *know* what ‘Being’ means. [...] we keep within an understanding of the ‘is’, though we are unable to fix conceptually what that ‘is’ signifies” (Heidegger 1962. 25). But without a proper understanding of the meaning of Being, how can we do philosophy? Not knowing what Being means will necessarily have an impact on philosophers (or indeed anyone) and their conception of anything because “basic concepts determine the way in which we get an understanding beforehand of subject-matter [...] [and] all positive investigation is guided by this understanding” (Heidegger 1962. 30). The concepts a subject works with guides how inquiry within it takes place and provides a framework for it. In philosophy, terms like ‘mind’ and ‘body’, and our understanding of them, provide a realm in which certain questioning and inquiry can take place. But ‘Being’, the most fundamental and universal concept, is implicated in every other – and we have no idea what it means. Just as you cannot teach a class without doing the necessary preparation, philosophers cannot expect to talk coherently about the nature of the mind, knowledge, goodness, truth, beauty or reality without working out or solving the question of the meaning of Being. In a nutshell, Heidegger's critique of the history of philosophy is that philosophers have been trying to run before they can walk. They have not done the requisite preparatory work, and their level of analysis was not fundamental enough for doing the things they wanted to do.

In the introduction to *Being and Time* (Heidegger 1962. 21–63), Heidegger lays out the first step in his program for overcoming this error – an “analytic” of

the being of what he calls *Dasein*. When we think about what Being is, we realise that Being is always the being *of* an entity – nothing can be if it has no Being. So if we are to find out what Being is, we should look to entities and analyse them. But not just any entity will do: it's hard to imagine what a rock or a table could tell us about the meaning of Being, beyond the fact that for something to exist it must have Being. But there is a special type of entity that would give us a better clue, “an entity which does not just occur among other entities” (Heidegger 1962. 32) like tables or rocks do, but which has a different, unique kind of Being that other entities do not. This type of entity does not just exist, but has a relationship of concern towards its existence. Its being “is an *issue* for it” (Heidegger 1962. 32), something that concerns it, something it must deal with. It is the only entity that we know of that can raise the question of the meaning of Being, and which has “certain ways of behaving that are constitutive for our inquiry” (BT 26) into it. To inquire into anything, there must be an inquirer that is capable of certain things, like “looking at something, understanding and conceiving it, choosing” (Heidegger 1962. 26) – but surely also asking and answering questions. The entity that Heidegger is describing, “those particular entities which we, the inquirers, are ourselves” (Heidegger 1962. 26–27) he calls *Dasein*. The question of the meaning of Being, the inquiry into it and the potential answering of it, is something that arises from and is made possible by the Being of *Dasein*. In pursuing this question, therefore, it is fundamental that we know what this being consists in such that we can, from out of this being, ask, understand and perhaps answer our guiding question. In short, “to work out the question of Being adequately, we must make an entity – the inquirer – transparent in his own Being” (Heidegger 1962. 27). To understand the question of the meaning of Being, we must first understand ourselves.<sup>5</sup>

The failure to conduct the right kind of analysis of our way of Being, specifically to prepare for the grounding-question, is what philosophers have neglected to do historically. According to Heidegger, this is the requisite preparatory work they have not done. He even at times calls his proposed analytic a “preparatory fundamental analysis” (Heidegger 1962. 65) of *Dasein*. ‘Preparatory’ not just because it prepares us for answering the grounding-question, or because the analysis will later be deepened in terms of time (Heidegger 1962. Division 2), but because once complete, this ontological analysis will prepare us for conduct-

<sup>5</sup> It is worth noticing (though not directly relevant) just how Kantian a move this is, recalling Kant's list of the four questions that unite philosophy. “What can I know? What should I do? What may I hope?” (Kant 1998. 677.) Kant's insight is that these questions imply a fourth that must be answered first: what is the human being? What is the nature of the being doing the knowing, acting and hoping? This leads Kant to conceive of philosophy anthropologically, and of its most fundamental questions as anthropological in nature. Heidegger makes exactly the same move with respect to the question of Being and *Dasein*.

ing philosophy in an appropriate, grounded fashion. To this end, Heidegger will concertedly avoid traditional philosophical vocabulary. Terms like 'subject', 'consciousness', 'mind', 'person' or 'human being' have little to no place in the analytic of Dasein, except to be criticised. Dasein, on Heidegger's account, is conceived essentially differently and in a new way from previous philosophical conceptions of the human being, and the introduction of 'Dasein' as a term also functions as a kind of cleansing of the philosophical palate, freeing us from any preconceptions of the entity to be analysed.

Heidegger's conception of his project, and the way he carries it out, involves a clear division: anthropology studies human beings, fundamental ontology studies Dasein. The focus, method, subject and scope of these inquiries, according to Heidegger, are so different and removed from one another that to study Dasein is to do something essentially different from simply studying human beings. Fundamental ontology, therefore, is not philosophical anthropology. In what follows, I will take a specific aspect of the analytic of Dasein – its account of moods – and show that the division Heidegger draws between his project and philosophical anthropology does not hold.

### III. HEIDEGGER'S ACCOUNT OF MOODS

It is always worth bearing in mind when discussing the analytic of Dasein that its analysis is a phenomenological one: it tries to describe what it is like to experience being Dasein and elucidate the structures necessary for this kind of experience to arise. It is a constant invitation to compare its analysis with one's own experience – and moods are a particularly important part of our experience of our existence. From how Heidegger discusses moods, I think it is clear that what he is talking about would include what we might normally call emotions. Some of the moods he discusses include fear (Heidegger 1962. §30), which he conceives as having many variations that can be identified with other emotional states, like dread, terror, timidity and shyness (Heidegger 1962. 182). He famously discusses anxiety, in *Being and Time* and *What is Metaphysics?* especially, where he also mentions joy as another example of mood (Heidegger 1998. 87). In *The Fundamental Concepts of Metaphysics*, he analyses boredom as a particularly important mood akin to anxiety. Our focus here will be on the general account Heidegger gives of moods in *Being and Time*, but it is worth bearing in mind that his use of 'mood' is very broad, and encompasses a large variety of affective phenomena and, based on the moods he explicitly discusses, would quite plausibly include what we normally call 'emotions'. But given that Heidegger speaks of moods as affective states we can find ourselves in that determine our way of being disposed to the world at a given time, 'mood' is quite an apt word for what he is talking about.

Heidegger's analysis of moods speaks about them in terms of *Befindlichkeit* and *Stimmung*, which I have opted to translate as 'disposedness' and 'mood', respectively.<sup>6</sup> They are intimately related, and defined in terms of each other as follows:

What we indicate *ontologically* by the term '*disposedness*' is *ontically* the most familiar and everyday sort of thing; our *mood*, our being-attuned. [...] it is necessary to see this phenomenon as a fundamental *existentiale* (Heidegger 1962. 172).

Disposedness and mood, in Heidegger's framework, are examples of what he calls *existentiale* structures. These are phenomena of Dasein's existence that are universal, necessary and constitutive for it: they are present in every instance of Dasein, and play a part in making its existence the way it is. If they were to be taken out of Dasein's existence, it would be a different entity entirely. 'Disposedness' refers to Dasein's capacity to be disposed, or 'attuned' some way towards its existence, in such a way that its existence and the things in it matter to it. Dasein is *always* disposed like this in some way towards its existence, and there are many ways we can be so disposed. The different ways Dasein can be disposed to its existence, the various possible manifestations of disposedness, Heidegger identifies as moods. In saying that moods are an *existentiale* structure, Heidegger is simply saying that all Daseins have moods, that moods play a part in structuring every Dasein's existence, and if we took away our moods we would no longer be the same entity. Without moods, our experience would be fundamentally different. Moods, being the manifestations of our disposedness, are what allow us to be disposed to the world and our existence in various ways. But how exactly is it that moods work, and how do they let us be disposed in different ways to our existence? Following Heidegger's answer to this question will allow us to view, in detail, Heidegger's most important claims about moods, and how they can be understood as contributing to the project of philosophical anthropology.

The aspects of Heidegger's account I will be focussing on are the following:

- (1) Moods are *fundamentally disclosive*.
- (2) Moods disclose our *thrownness*.
- (3) Certain moods disclose our being-in-the-world as a whole.<sup>7</sup>

<sup>6</sup> There is debate over how these terms (especially *Befindlichkeit*) should be translated, but I will not go into that here. Debating the translation of these terms is not necessary to understand the general claims Heidegger makes about the phenomena he is discussing, and would distract from the point. For an indication of the depth of disagreement on this, William Blattner has provided a helpful compilation of various ways that Heidegger's terms have been translated (<https://faculty.georgetown.edu/blattnew/heid/Heidegger-jargon.html>).

<sup>7</sup> Heidegger also claims that moods let things matter to us, but since this is perhaps unconvincing, not as interesting as the first three claims and not necessary to consider for the purpose of my argument, I will leave it aside.

Heidegger casts moods as fundamentally disclosive, meaning that their basic function is to disclose things, to reveal information to us about the world and things in it, and this process happens on a more fundamental level than cognition, reason and knowing.<sup>8</sup> A simple way to understand how moods can be fundamentally disclosive is to consider perhaps the most intuitive thing that moods reveal to us, expressed by Heidegger as follows: “a mood makes manifest ‘how one is, and how one is faring’.” (Heidegger 1962. 173) Moods are states we find ourselves in that let us know how we are doing, what our general affective state is – ‘I am happy’, ‘I am bored’, etc. and the mood I am in determines how I experience my existence and the things I encounter in it, at that time. Moods give us an emotional context through which to experience our existence, affecting our perception of the things in it. If I am happy, I will be more likely to experience certain things (or people) as joyful, or uplifting, rather than as annoying, or angering. In their disclosure of ‘how we are doing’, moods provide an affective setting for our experience to take place in, which alters our experience of our world depending on what mood we are in, and constitutes an integral part of our making sense of our environment.

This is where the idea of “fundamental disclosure” really comes into play, because moods do not just disclose “how we are doing” – *everything* in our experience is disclosed to us through some mood or another since we are never un-mooded, and this disclosure takes place on a more fundamental level than reason, cognition and knowing. These things have often been taken to be the primary way we engage with and relate to the world, but on Heidegger’s account moods operate “prior” to these things and “beyond their range of disclosure” (Heidegger 1962. 175). This is because even before we can reason, deduce, know (etc.), we have moods, and we still make some kind of meaningful sense of our world, we just do so in a non-linguistic and pre-reflective way. Babies, even though they cannot talk, or think philosophically, still make some kind of sense of the things around them, and things are still disclosed as having some sort of significance, even if this significance is diminished in comparison to what it might be for an adult. Babies still have moods, and these moods still disclose things as significant to them, even if the level of disclosure only occurs in terms of pleasantness or unpleasantness, or of liking or disliking a particular toy, thing, or food, etc. But this is true of all moods no matter what age we are – they dis-

<sup>8</sup> This is perhaps beside the point here, but this is much that could be said about how Heidegger’s account of moods overturns some traditional conceptions of human affectivity, where moods are understood as “internal, subjective mental states [...] caused by one’s external situation” (Elpidorou–Freeman 2015. 664). Furthermore, Heidegger’s account of moods is a fundamental rejection of a certain way of conceiving our connection to the world that many philosophers have adopted over the centuries, where the primary way we connect with our world is through reason, rationality and knowing. For Heidegger, these things are made possible by a more fundamental and pre-reflective engagement with our environment, of which moods are an integral part.



close things to us, and do so on a level different to that of rational disclosure. Think again of how moods disclose ‘how we are doing’ – this is not something we reason ourselves into, or find out by way of thinking. It is something that is revealed to us experientially, non-linguistically, and unreflectively – we find ourselves in a state. This is what Heidegger means when he says that moods are fundamentally disclosive: they reveal information to us about ourselves, our worlds and the things in them on an existentially prior level to rationality, and play a part in enabling rationality. Without the pre-reflective familiarity with the world that we have before we can reason, reason would not be possible. Moods are an integral part of our most fundamental way of making sense of the world, and play an important role in how things are disclosed to us as meaningful in the first place. We perceive and understand things, they have a certain place in our world, and moods partly constitute the process by which we apprehend them, and disclose them as meaningful.

But moods disclose other things apart from our general affective state, or things in the world. They are also said by Heidegger to “disclose Dasein in its thrownness” (Heidegger 1962. 175) and, in certain cases, disclose insight about our being-in-the-world as such. Heidegger’s notion of ‘thrownness’ is perhaps best approached with a quote from Kierkegaard, talking about life: “no one asks when one wants to come in; no one asks when one wants to go out” (Kierkegaard 1987. 26). Our existence is something that is forced upon us, unchosen, but we have to deal with it – we are “thrown” into the world and burdened with the responsibility of existing. We do not decide to be born, and our lives are tinged by the fact that we know we will one day die – how our existence begins and ends is something we have no control over, and this is also the case with many things along the way. This renders our existence, in some sense, quite strange, difficult, or even absurd, and it is this unchosen, disquieting aspect of our predicament that Heidegger refers to as Dasein’s “thrownness [...] [which] is meant to suggest the facticity of its being delivered over” (Heidegger 1962. 174) to its existence.

Interestingly, Heidegger claims that it is through moods that our thrownness is revealed to us, and gives at least two reasons for this. We have already seen how moods are a constitutive part of how we make sense of the world in our most primal way. The mood is what “brings Dasein before the ‘that-it-is of its ‘there’” (Heidegger 1962. 175), allowing us to make sense of the world and the things in it. If our thrownness is something that is disclosed to us, moods must therefore play a part in its disclosure by definition – if thrownness is a part of our existence that we can encounter and make meaningful sense of, our moods must at least partially constitute the process of its disclosure to us. The second reason we can draw from Heidegger as to why moods disclose thrownness has more to do with the experience of being in a mood: there is a similarity between thrownness and moods as we experience them. Just as we do not choose to be



thrown into the world, we do not really choose to be in a particular mood or another: “a mood assails us” (Heidegger 1962. 176). I do not choose to enter a bad mood, nor do I choose to be in a happy one – there is no switch that we can turn on and off when it comes to our mood. We can do certain things that might affect it, such as putting on a depressing film, or eating a delicious meal, but we cannot directly choose to change our mood, and force our body and mind to adapt to our preferred mood, or preferred degree of mood. Even if I know that putting on Roman Polanski’s *The Pianist* will make me sad, I have no control over how sad I will be, or which particular variety or degree of sadness I will feel – this is beyond my control, and I am ‘thrown’ into it, like I am thrown into my existence.

In certain special cases of mood, existential insights about our being-in-the-world as such are disclosed to us. Heidegger famously identified anxiety (*Angst*) as an example of this.<sup>9</sup> Rather than disclosing particular items or sets of items in the world, anxiety (or at least a specific variety of it) is a rare case of mood that can disclose the totality of our world to us. This is because the experience involved in them is one of a total change of the structure of our world as we normally experience it. Normally, our worlds and everything in them are invested with and seen in the light of significance, with our moods disclosing the objects in them as significant in a particular way. When afraid, for example, we are always afraid of particular things. These things are disclosed as having a meaning, and being worthy of fear. But “that in the face of which one is anxious is completely indefinite”, not *an object* at all but “*the world as such*” (Heidegger 1962. 231). Heideggerian anxiety is a phenomenon where we completely withdraw from our usual world. We become so paralysed by the weight of our anxiety that we temporarily fail to make sense of the world as we normally do - things become *insignificant* to us. In such a state, we view the world from a completely different perspective, which tells us what it is otherwise always like to be in one. Elsewhere, Heidegger calls this a ‘telling refusal’ (Heidegger 1995. 137) – in temporarily and totally refusing itself, the nature of the world announces itself more forcefully.

This is why anxiety is an encounter with “Dasein’s *innermost freedom*” (Heidegger 1995. 136). What we are anxious about in such states is not a particular thing, or set of things, but the weight of our freedom as such. We become so overwhelmed by our possibilities that we are left dumbstruck, paralysed before our world and our freedom. But in such a state, where the usual character of the world – as structured by significance – announces itself, we are afforded a unique insight into our role within it. We are the beings that are responsible for the creation, maintenance and inhabiting of the very significance of our worlds. Who wouldn’t feel anxious upon realising this for the first time? It is an unsettling experience, but one that forcibly confronts Dasein with its “*Being-free* for

<sup>9</sup> Less famously but just as interestingly, he also identifies boredom as one, but I will focus on anxiety here for considerations of space.

the freedom of choosing itself and taking hold of itself” (Heidegger 1962. 232). Anxiety functions as a powerful existential catalyst for being able authentically seize hold of our lives, and appreciate more forcefully and authentically our role as free creators and keepers of significance.

#### IV. HEIDEGGER'S PHILOSOPHICAL ANTHROPOLOGY OF MOODS

We have now seen some of the most important claims Heidegger makes about Dasein's moods, which puts us in a good position to question Heidegger's claim that his work is not philosophical anthropology. What is important to bear in mind is Heidegger's conception of his own project as an analysis of Dasein, and the incontestable fact that human existence is a case of Dasein. When Heidegger analyses Dasein, he is doing so from his own perspective of his existence as a human being. The structures he identifies are ones that structure his own existence and human existence in general. He may be analysing human beings only insofar as they are cases of Dasein, but this is nonetheless true. *Anything that applies truly to Dasein applies truly to human beings, because human beings are Dasein.* There is nothing that can be said truly of Dasein's existence that is not true of human existence, and discovering true things about the nature and structure of human existence is precisely the aim of a philosophical anthropology. If Heidegger is successful in disclosing anything true about Dasein at any point, he has engaged in philosophical anthropology, even if he is not keen on admitting it.

With this in mind, let's consider some of the claims Heidegger makes about moods. This list is not exhaustive, and there is room for debate about whether or not these are all separate claims, but it is sufficient for our current purposes.

- (1) Moods are *fundamentally disclosive*: they are part of the process through which we make sense of the world and disclose the things in it as meaningful on a pre-rational, pre-linguistic level.
- (2) Moods disclose 'how we are doing', providing an affective context to experience our existence through at a particular time, and be disposed a certain way towards it.
- (3) Without the fundamental disclosure of moods, the forms of disclosure associated with reasoning and knowing would not be possible.
- (4) Moods constitute the different ways we can be disposed to our existence.
- (5) Moods disclose our 'thrownness'.
- (6) Certain varieties of mood, like anxiety and boredom, reveal information not about specific things in the world, but about our being-in-the-world as such, thereby conveying insights to us about what it means to be the kind of entity we are.
- (7) These insights concern our freedom and role in the creation and maintenance of the significance of our worlds.

These are claims that, if true, would apply generally to human beings and their experience of moods, stating true things about the nature and structure of human existence – they therefore would be appropriately classed as philosophical-anthropological. Perhaps they are not all true, but even if they are not, Heidegger's work still gives us (at the very least) an attempt at philosophical anthropology, because he is reflecting on his experience and attempting to deduce general truths about it that would apply to all human beings. These are all claims that speak to what it is like to be a human being, to experience the kind of existence human beings have, and would contribute to an understanding of the structures of human existence.

We do not necessarily need these claims to be true to make the argument that Heidegger is doing a kind of philosophical anthropology, we just need to examine the kind of claims he is making and what kind of project he is engaged in. Based on what I have said, I think it is evident that Heidegger is mistaken to be distancing himself so sharply from philosophical anthropology, given the powerful potential his work has for contributing to it, and for benefitting our understanding of the human experience. But the stronger argument is certainly possible to make: Heidegger does manage to state true things about moods, and therefore manages to successfully contribute to philosophical anthropology. I will focus on one in closing my argument. Surely one of the achievements of *Being and Time* (and one of the things it spends most of its time doing) is showing that our most basic way of engaging with the world is not one of knowing, reasoning or rationality. We in fact have a deeper, pre-reflective, more primal way of being in the world that makes these former ways of being possible. Moods would quite plausibly be an integral part of this primal sense-making process, and a constitutive structure for the way we exist. There is therefore much value in Heidegger's working-out of the idea that moods are 'fundamentally disclosive', and perhaps a lot of truth. But if there is, it is truth that contributes to philosophical anthropology, and Heidegger's analysis of Dasein as a whole should be read as such a contribution. It is at its most productive when understood this way, and he gave us no convincing reasons for its essentially differing from philosophical anthropology.

## REFERENCES

- Benedict, Ruth 2005. *The Chrysanthemum and the Sword: Patterns of Japanese Culture*. Boston – New York, Mariner Books, Houghton Mifflin.
- Blattner, William Translations of Heidegger's Jargon.  
<https://faculty.georgetown.edu/blatnew/heid/Heidegger-jargon.html>
- Elpidorou, Andres – Lauren Freeman 2015. Affectivity in Heidegger I: Moods and Emotions in 'Being and Time'. *Philosophy Compass*. 10/10. 661–671.
- Heidegger, Martin 1962. *Being and Time*. Trans. John Macquarrie and Edward Robinson. London, Blackwell.

- Heidegger, Martin 1973. *Kant and the Problem of Metaphysics*. Trans. Richard Taft. Bloomington–Indianapolis, Indiana University Press.
- Heidegger, Martin 1995. *The Fundamental Concepts of Metaphysics: World, Finitude, Solitude*. Trans. William McNeill and Nicholas Walker. Bloomington–Indianapolis, Indiana University Press.
- Heidegger, Martin 1998. *What is Metaphysics?* (collected in *Basic Writings*). Trans. David Farrell Krell. Cambridge, Cambridge University Press.
- Heidegger, Martin 1999. *Contributions to Philosophy (From Enowning)*. Trans. Parvis Emad and Kenneth Maly. Bloomington–Indianapolis, Indiana University Press.
- Heidegger, Martin 2016. *Ponderings II–VI: Black Notebooks 1931–1938*. Trans. Richard Rojcewicz. Bloomington, Indiana University Press.
- Heidegger, Martin 2017a. *Ponderings VII–XI: Black Notebooks 1938–1939*. Trans. Richard Rojcewicz. Bloomington, Indiana University Press.
- Heidegger, Martin 2017b. *Ponderings XII–XV: Black Notebooks 1939–1941*. Trans. Richard Rojcewicz. Bloomington. Indiana University Press.
- Kant, Immanuel 1998. *Critique of Pure Reason*. Trans. Paul Guyer and Allen W. Wood. Cambridge, Cambridge University Press.
- Kierkegaard, Soren 1987. *Either/Or*. Trans. Howard V. Hong and Edna H. Hong. New Jersey, Princeton University Press.
- Kluckhohn, Clyde 1949. *Mirror For Man*. New York, Fawcett.
- Kottak, Conrad Phillip 1997. *Anthropology: The Exploration of Human Diversity*. New York, McGraw-Hill.

PHILIPPE CABESTAN

## Bad Faith *versus* Unconscious: a Credible Alternative?

It's certainly not too late, after more than one century of psychoanalysis, to evaluate again the main hypothesis on which Freud's research is based, I mean the hypothesis that the major part of our psychological life is unconscious. Everybody knows, of course, Freud's comparison between the psyche and an iceberg, according to which ninety percent of the psyche is unconscious.

Around forty years ago, when I was a student, this hypothesis was so obvious that it seemed there was no way to criticize it. Just like we "knew" from Marxism that history was the history of class struggles, we "knew" from depth psychology that we were more or less neurotic and that our lives were dominated by unconscious conflicts, which were deep-rooted in our childhood. In other words, we were absolutely convinced that Freud, as he wrote himself, had inflicted, after Copernicus, after Darwin, a third blow to the universal self-love of the humanity. This blow, which is psychological in nature, would be the discovery that "the ego is not master in its own house" (Freud 1917/1955. 135. et sq.).

But now, in 2019, it has to be said that Freud's way of thinking has already lost a part of its credit, especially among the youth and even, perhaps, among the psychoanalysts themselves. However, it's certainly not a reason to declare that the hypothesis of the unconscious is irrelevant, null and void. Therefore, I would like to examine whether and how it's possible to keep Freud's hypothesis. My main reference will be obviously Jean-Paul Sartre, but because it seems to me that Sartre's criticism of psychoanalysis has to be understood as an extension of Heidegger's ontological point of view, I would like to begin with Heidegger's criticism of Freud's hypothesis. Then, I'll consider Sartre's concept of bad faith and develop the idea of a negative psychology.

## I. MARTIN HEIDEGGER, MEDARD BOSS AND THE PRINCIPLE OF SUFFICIENT REASON

The loss of credit of psychoanalysis doesn't come as much of a surprise, if we take into account the fact that psychoanalysis has been criticised for a long time and since the very beginning by various philosophers. We think of Karl Jaspers – and I'm glad to pay tribute to his memory – but also of Ludwig Wittgenstein, Karl Popper, Adolf Grünbaum, Alasdair MacIntyre, Jürgen Habermas and also, if we limit ourselves to French thinkers, of Jean-Paul Sartre, Emmanuel Levinas, Michel Foucault, Gilles Deleuze, Vincent Descombes, and so on (our list is obviously not exhaustive). But if I had to choose among all these critics, with whose works – honestly speaking – I'm more or less familiar, it's certainly Heidegger and its point of view I would first select, as the most relevant one. In fact, Heidegger knew very little about psychoanalysis and was apparently not concerned by Freud's research. He had however the opportunity to speak about Freud's work when he met the Swiss psychiatrist Medard Boss after the Second World War, and also during a seminar organised both by Heidegger and Boss in Zollikon, at the home of Medard Boss, between 1959–1969. We have the protocols of the seminars, which can be completed with Medard Boss's own writing about psychoanalysis. (Heidegger 2001; Boss 1963; Escoubas 1992.)

Heidegger's review is based on an ontological argument, according to which the categories we use to describe natural phenomena don't fit human phenomena. For natural phenomena belong to one form of being whereas human phenomena belong to another. It means you can't think and speak of a human behaviour as if it was a physical or chemical process. As a result, human sciences have to ban from their vocabulary such words as "energy", "force", "cause and effect" which are used in the fields of modern natural sciences. From this point of view, it's easy to understand why Heidegger refuses Freud's hypothesis. As a matter of fact, Freud's metapsychology is full of this kind of concepts and, as we know, Freud identifies *libido* with a sexual "energy". But there is an other critical argument, a bit more difficult to explain: when Freud, in a very famous text, wants to justify the concept of unconscious, he writes: "the data of consciousness have a very large number of gaps in them; both in healthy and in sick people psychical acts often occur which can be explained only by presupposing other acts, of which, nevertheless, consciousness affords no evidence" (Freud 1914–1916. 165). We have therefore to fill in the blanks with the unconscious acts in order to establish the full text. But if we understand this statement with Heidegger, it follows that Freud's way of thinking is based on a kind of ontological prejudice, which comes from modern natural sciences, that's the idea of a causal connexion and causal explanation without gaps or blanks. From this point of view the first root of Freud's Hypothesis is the modern principle of causality. Since he presumed an unbroken causal connexion in the psychical life, Freud

imagines unconscious acts. The door is now open for a more elaborate hypothesis of the unconscious itself as the main part of the mental life. (Heidegger 2001. *Conversation with Medard Boss in 1965*).

## II. FREUD'S AMBIGUOUS CONCEPTION OF THE UNCONSCIOUS AND SARTRE'S CONCEPT OF BAD FAITH

It seems to me that Sartre's criticism of psychoanalysis has to be understood as an extension of Heidegger's ontological point of view. In other words, Sartre's criticism is based again on an ontological argument and on the conviction that human mode of being excludes such a *thing* – and let me stress this last word – such a *thing* as the unconscious. In *Being and Nothingness* (Sartre 1994), as we know, Sartre opposes two different modes of being: on the one hand, the thing as in-itself (*l'en-soi*): for example, a stone or a cigarette is an in-itself, and, on the other hand, the conscious as a for-itself (*le pour-soi*): for instance, a desire or a feeling, as given to the consciousness, shares necessarily the mode of being of the consciousness. From this point of view, it's meaningless to speak of an unconscious desire, of an unconscious guilt feeling and so on. And not, as is often said, because the subject is transparent to itself – such an idea never occurs by Sartre who always distinguishes between the consciousness of self and the knowledge of self.<sup>1</sup> But because being conscious is both a quality of a lived process and a mode of being, so that desires or feelings as lived processes share therefore the mode of being of what is for-itself. For instance, to imagine something is nothing else but an act of the consciousness and has necessarily an intentional and temporal form or structure. If not, the imagination is then an impossible and absurd mix of in-itself and for-itself, which is supposed to subsist in the unconscious part of the psyche like a picture in a gallery.

One can object that these considerations don't concern Freud's hypothesis, in so far as the Freudian unconscious can't be assimilated to a thing in-itself. From this point of view, the comparison of the psyche with an iceberg is just a comparison and must not be taken literally. In fact, Freud's description of the unconscious is relatively ambiguous. Paul Ricœur is certainly right when he says, in his essay about Freud, that "Freud's writings present themselves as a mixed or even ambiguous discourse, which at times states conflict of forces subject to an energetics, at times relations of meaning subject to a hermeneutics" (Ricœur 1970. 65, 395). As a result, it's possible to read from Freud himself a description of the unconscious that substitutes the economic language for an intentional one. For example, when it comes to repression, Freud compares this pheno-

<sup>1</sup> Borrowing the expression to Maurice Barrès, Sartre speaks of a "mystery in broad daylight" (Sartre 1994. 571).

menon with the intentional action of a watchman.<sup>2</sup> Of course and again, it's only a comparison and Freud underlines himself its unscientific character<sup>3</sup>. But this kind of presentation is nevertheless meaningful of a difficulty that Sartre points out in *Being and Nothingness*: how could a blind force repress unpleasant drives or impulses without being aware of their unpleasant character for the ego? In other words, which I borrow from Sartre, "if we abandon all the metaphors representing the repression as the impact of blind forces, we are compelled to admit that the censor must choose and in order to choose must be aware of so doing" (Sartre 1994. 52).

Everybody knows, I suppose, Sartre's alternative proposal to Freud's concept of the unconscious, which can be summed up in the concept of bad faith or, as it has been proposed, in the concept of self-deception (Soll 1981. 584). Quite close to Sartre on this matter, Merleau-Ponty, for its part, speaks of a "metaphysical hypocrisy" (Merleau-Ponty 1985. 190). From this point of view, the so-called repression of an idea is nothing but the act of lying to oneself and the real question is then to understand how it's possible for a subject to lie to himself. For Sartre is convinced that the hysteric as a neurotic is aware of what he doesn't want to be aware of and, as long as he tries to escape from it, the hysteric is necessarily aware of what he tries to escape from. He would not be trying otherwise to escape from it. In *Being and Nothingness*, Sartre quotes Freud's dissident follower, Wilhelm Stekel who states: "Every time that I have been able to carry my investigations far enough, I have established that the crux of the psychosis was conscious" (Sartre 1994. 54). And the same goes for the homosexual who refuses with all his strength to consider himself as a homosexual, even if he (or she) recognises to have had sexual relationships with people of the same gender (Sartre 1994. 54, 63).<sup>4</sup>

In this way of thinking, it's worth examining MacIntyre's plea, after Wittgenstein, for an adverbial conception of the unconscious (MacIntyre 1958. 77; 1984. 71). According to MacIntyre, there are two ways of using the word 'unconscious': either as an adjective and as an adverb or as a noun and as a substantive. In the former case, the adjective unconscious enables us to describe human pheno-

<sup>2</sup> The watchman stands between the drawing room and a large entrance hall, and doesn't admit into the drawing room the impulses which displease him (Freud 1916–1917. Ch. 19, *Resistance and repression*).

<sup>3</sup> « On peut donc dire que la conscience vire les désirs indésirables à la manière dont un enseignant met à la porte les élèves qui perturbent la classe » (Bernet 2013. 343).

<sup>4</sup> One can object that Sartre misses the point. What is actually at stake, according to Freud, is not the homosexuality of one who is aware of his sexual orientation but deny to be homosexual. But it is, as well as in Leonardo da Vinci's case studied by Freud, the unconscious drives of a man who preserves his love for his mother by repressing it and who seems to pursue boys and to be their lover while he is in reality running away from other women who might cause him to be unfaithful to his mother. But it's easy to conceive Sartre's reply: Leonardo can run away from other girls and pursue boys only if he is aware of his so called unconscious love for his mother. The concept of bad faith is back again. (Freud: *Leonardo...* Ch. 3.)



mena like dreams, jokes, neurotic symptoms, traumas, and so on; in the latter, as noun, unconscious pretends to explain and to give the cause of these kinds of phenomena. In others words, according to MacIntyre (and, of course, Freud), a great deal of human phenomena express an unconscious intention and we have therefore to extend our concept of intention; but we have no evidence and no need of the unconscious as a part of the psyche. Let's take the case of "a man involved in an unhappy love affair who tells his friend that he intends to break free from it, but who continues to see the girl and to send her gifts". MacIntyre asks: "what are we to say his intentions in fact are?" Two interpretations are at least possible. First, the man lies to himself, he acts in bad faith; second, the man is divided between two conflicting intentions: his conscious intention is to leave the girl, the unconscious intention, only expressed by his behaviour, is to stay along with her (MacIntyre 1958. 84).

Let's take another example: Chimène's understatement (*litote*): "Go I do not disdain you" (*Va, je ne te hais point*). The sentence comes from Pierre Corneille's famous tragicomedy, *Le Cid*. As it is well-known at least in France, Rodrigue and Chimène were in love; but the marriage became suddenly impossible for Rodrigue killed Chimène's father in a duel; Chimène states that she hates Rodrigue; but, as Rodrigue offers Chimène to kill him, Chimène finally declares: "Go I do not disdain you". This sentence is interpreted by Vincent Descombes as if Chimène was unconsciously (adverb) expressing her love. We would have therefore to distinguish between the intentional anger which Chimène declares and the intentional love which is declared by Chimène's sentence but of which Chimène is unconscious. But a much simpler interpretation is also possible without such a hypothesis like unconscious intention: "Go I do not disdain you" expresses very well the evolution of her emotional life and her current state of mood.<sup>5</sup> In others words, because Rodrigue killed her father, Chimène is no more in love with Rodrigue but at this point she no more suffers because of hate. It follows from this short analysis that the concept of bad faith can quite well clarify behaviours, whose elucidation seems demand unconscious intention.

### III. THE UNCONSCIOUS AS CONCEPT OF A NEGATIVE PSYCHOLOGY

It's time to raise the ultimate question, which is the real guideline of this paper: if we give up Freud's hypothesis of a mental unconscious as a thing in-itself, can we be satisfied in any cases with the concept of bad faith? In other words, I would like to develop the idea that there are some phenomena, which call an-

<sup>5</sup> P. Corneille, *Le Cid*, acte III, scène 4, vers 963. Descombes 1984; 1977/ 2004.

other interpretation. The study of these phenomena leads perhaps to a renewed concept of the unconscious. In this context, I'm going to consider the link that Freud introduces between homosexuality and the unconscious.

In his essay about Leonardo da Vinci, Freud expresses a general explanation of homosexuality as a result of an unconscious love for one's mother and of a psycho-sexual development. But to tell the truth, the most important part of Freud's explanation is not the explanation itself but the following part, when Freud comments on his own explanation. For Freud, regarding Leonardo's homosexuality is fully aware of the much too general character of the incestuous love. So, he adds this cautious remark: "we cannot reject the supposed cooperation of unknown constitutional factors" (Freud: *Leonardo* ... Ch. 3). Of course, as a scientist, as a follower of the Enlightenment, as a hyper-rationalist, Freud is convinced that these constitutional factors are not destined to remain unknown. But, and here is the thesis I would like to defend, it's possible to give another conception of these "unknown constitutional factors", if we put aside the principle of reason as well as the way of thinking which is based on it.

So influential is this principle upon our mind that we cannot help but think under its control. However, when Jean-Paul Sartre, in his essay on Jean Genet, tries to understand why Jean Genet became a thief, a poet and a homosexual, his interpretation stands actually beyond the principle of reason, in so far as Sartre relates homosexuality to Jean Genet's own tragic situation and to a free choice. From this point of view, homosexuality has nothing to do with constitutional factors but, like the genius, is "the way out some one invents in desperate cases" (Sartre 1952. quatrième de couverture). However, even if Sartre's thesis deserves our attention, it has to be said that this appeal to freedom doesn't fit the lived experience. But there is, of course, another way of understanding the phenomenon of sexual orientation. If we admit the limit of all kinds of explanations, we can understand these constitutional factors as an *innate* predisposition. Innate: the adjective sounds perhaps strange. We speak for instance of an innate gift for music or for mathematics, that is to say of an unexplainable and unacquired feature of the character. In this same way of thinking, when the German psychiatrist Hubertus Tellenbach, in his major work on melancholia, shapes the *typus melancholicus* as a personality vulnerable to melancholia, he carefully avoids asking why this one or that one has such a premorbid profile (Tellenbach 1961).

As regards homosexuality (or heterosexuality), the adjective "innate" means on the one hand that you don't choose your sexual orientation and on the other hand that besides all explanations –more or less credible – from the family, the education, the personal story, the society and its heteronormativity, and so on, remains a dark, unknown and unknowable side of the personality. Far from meaning that the ego is master in its own house, it means the opposite, that its existence comes actually from deeper than itself, than its history, from a depth

or a dark side which cannot be enlightened.<sup>6</sup> Therefore, I would like to argue in favour of a negative psychopathology. That is to say, if I may make such an analogy, just as we know from negative theology that it is impossible to say anything positive about God, we ought to know from negative psychology that we can't unlock the enigmatic character of a personality. This issue has nothing to do with the ethics but with our senseless claim to know totally a person. It follows, given the mode of being of the subject, that it's possible to state that the unconscious is neither a result of a past and repressed event, nor a thing, a device or a machine. And, if we don't want to deceive ourselves with unjustified speculations, we have to say, but according to a phenomenological point of view, "whereof one cannot speak whereof one must be silent" (Wittgenstein 1999. 7).

#### IV. CONCLUSION

My starting point was the opposition of Freud's unconscious and Sartre's bad faith. The question was: does Sartre's bad faith constitute a credible alternative to Freud's unconscious? Of course, it does *and* it doesn't. It does, for a lot of behaviours are understandable without the obscure hypothesis of the unconscious, which offers then no more light. It doesn't, if we take into account the story of the subject in his time. The roots of this unconscious would be the past of the subject and its ontological mode would be the enigmatic one of the past in so far as it runs the present subject. But, regarding what Freud himself calls "constitutional factors", there is another and obvious mode of the unconscious.<sup>7</sup> These "constitutional factors" represent the innate part of the human being, which belongs to its nature and which also runs its behaviour. It can be a gift or a burden, which in any cases the subject doesn't choose and has to assume. Of course, this latter mode of the unconscious is quite different from the former. It has no immediate links with the past of the subject and belongs to what we can call with Heidegger the *facticity* of its existence. One can feel disappointed in this conception since there are no more stories to imagine and to tell about it, lying on the couch. But if we are able to back away from the principle of reason, we are ready to admit the limit of a kind of psychological inquiry, which tries to explain what is unexplainable and must remain unexplainable.

<sup>6</sup> From this point of view I agree with Rudolf Bernet, when he writes in his last book about the subject as the subject of the unconscious: "For the most fundamental question is not whether the ego is an object or a subject but whether the subject of the desire [...] is the producer of his desire or just an effect of fantasies which overwhelm him and come from deeper than himself" (Bernet 2013. 336).

<sup>7</sup> With W. Blankenburg and Arthur Tatossian, it's possible to consider a third mode of unconscious that is the transcendental unconscious as the *a priori* condition of the natural attitude. Belongs, for instance to this mode of unconscious the natural self-evidence that is unfortunately missing by schizophrenia. (Blankenburg 1991; Tatossian 1997; 2020).

## REFERENCES

- Bernet, Rudolf. 2013. *Force-pulsion-désir : une autre philosophie de la psychanalyse*. Paris, Vrin.
- Blankenburg, Wolfgang 1991. *La perte de l'évidence naturelle*. Trad. fr. J.-M. Azorin et Y. Totoyan. Paris, PUF.
- Boss, Medard 1963. *Psychoanalysis and Daseinsanalysis*. Trans. Ludwig B. Lefevre. New York, Basic Books.
- Descombes, Vincent 1977/2004. *L'inconscient malgré lui*. (Folio essai). Paris, Gallimard.
- Descombes, Vincent 1984. L'inconscient adverbial. *Critique*. 449. 775–795.
- Escoubas, Eliáne 1992. « La fatale différence ». Ontologie fondamentale et archéologie de la psyche : Heidegger et Freud. In J.-F. Courtine (éd.) *Figures de la subjectivité. Approches phénoménologiques et psychiatriques*. Paris, Editions du CNRS. 147–164.
- Freud, Sigmund 1910/1957. *Leonardo da Vinci and a Memory of His Childhood*. Trans. A. Tyson. *The Standard Edition of the complete psychological works of Sigmund Freud*. Vol. 11.
- Freud, Sigmund 1914–1916/1966. *The Unconscious*. The Standard Edition of the complete psychological works of Sigmund Freud. Vol. 14.
- Freud, Sigmund 1916–1917/1963. Introductory Lectures on Psycho-analysis. In *The Standard Edition of the complete psychological works of Sigmund Freud*. Vol. 16.
- Freud, Sigmund 1917/1958. A Difficulty in the Path to Psycho-analysis. In *The Standard Edition of the complete psychological works of Sigmund Freud*. Trans. J. Strachey. The Hogarth Press, Vol. 17.
- Heidegger, Martin 2001. *Zollikon Seminars: Protocols, Conversations, Letters*. Trans. F. Mayr. Chicago, Northwestern University Press.
- MacIntyre, Alasdair. C. 1958. *The unconscious. A conceptual analysis*. New York – London, Routledge.
- MacIntyre, Alasdair C. 1984. *L'inconscient. Analyse d'un concept*. Trad. fr. G. Nagler et P.-E. Dauzat. Paris, PUF.
- Merleau-Ponty, Maurice 1985. *Phénoménologie de la perception*. (Collection TEL). Paris, Gallimard.
- Ricœur, Paul 1970. *Freud and Philosophy: An Essay on Interpretation*. Translated by Denis Savage. New Haven, Yale University Press.
- Sartre, Jean-Paul 1952. *Saint Genet, comédien et martyr*. Paris, Gallimard.
- Sartre, Jean-Paul 1994. *Being and Nothingness*. Trans. Hazel E. Barnes. New York, Gramercy Books.
- Soll, Ivan 1981. Sartre's Rejection of the Freudian Unconscious. Paul A. Schilpp (ed.) *The Philosophy of Jean Paul Sartre*. (The Library of Living Philosophers; 16). La Salle, Ill, Open Court.
- Tatossian, Arthur 1997. Inconscient et phénoménologie. *Psychiatrie et phénoménologie, ETIM*. 143–145.
- Tatossian, Arthur. 2020. La phénoménologie existentielle et l'inconscient. Un point de vue psychiatrique. In *Œuvres complètes*. Paris, MJWF édition. Vol. 3. 7–28.
- Tellenbach, Hubertus 1961. *Melancholie: Problemgeschichte, Endogenität, Typologie, Pathogenese, Klinik*. Berlin–Heidelberg, Springer Verlag.
- Wittgenstein, Ludwig 1999. *Tractatus logico-philosophicus*. Trans. C. K. Ogden. Mineola – New York, Dover Publications.

## Alienation

In this paper I discuss three paradigmatic thinkers of alienation: Rousseau, Marx, and Lukács. I am going to focus on two major aspects in their work which are of interest for a project I named neoexistentialism. The first major aspect can be expressed with the question of what should be reappropriated in overcoming alienation? The second point concerns the question of how we experience our being alienated? Or put otherwise, what kind of emotions, moods, if any, are indicative of being alienated? This second issue is connected to the problem of whether unconscious alienation is possible or not? If possible, it should be explained how her own being alienated can be made accessible to the person in question.

In particular, I examine with regard to Rousseau the structure of what might be called a precursor conception of alienation. By this, I mean that alienation in a broad sense could be and has been understood in Rousseau as an analysis of “social pathologies” in the development of modern society. However, alienation in this sense has the structure of possession and subsequent disappropriation of man’s original constitution. If we take a closer look at Rousseau in the light of Marx’ more specific concept, it can be pointed out that there is a general structure of alienation that might be described with the possession – disappropriation – reappropriation formula. In Rousseau, I claim, we have a simplified version of alienation in the form of hypotheticalal possession – disappropriation.

The discussion of Rousseau already implies a look at Marx’ theory of alienation which I develop in the second, short part of my paper. Here I show that normative basis of alienation in the early Marx is the concept of man’s self-realization in the working process. The self-realization, in turn, takes place in a double movement of a prior objectification and a following re-appropriation.

The last part of my argumentation is dedicated to Lukács’s theory of reification in *History and Class Consciousness* where his contribution to the theory of alienation has often been seen in his concept of reification (*Verdinglichung*). I discuss Lukács’s critique of capitalist society with an eye on how the concept of reification partly carries on and partly modifies Marx’s conception of alienat-

ed labor. This part of the paper shows that even Lukács could not clarify how non-alienated conditions should be conceived – a problem recent descriptions of alienation (Hartmut Rosa, Rahel Jaeggi) could not solve either.

## I. ROUSSEAU ON ALIENATION

It is widely accepted that Rousseau gave one of the fiercest critiques of modern civilization and society in general. He has been often said to have accurately described or at least anticipated what later on was baptized alienation,<sup>1</sup> even if he himself made no use of the term. With regard to this concept of alienation, Rousseau's *Second Discourse* on inequality might serve as a point of reference. In this text, he essentially claims that (1) human beings under conditions of civilization, while seemingly free, are in fact enslaved by their mutual social relations, and (2) that this slavery to one another was brought about by a socio-cultural development leading to a loss of authenticity. Alienation in a broad sense could be and has been understood in Rousseau as an analysis of "social pathologies" in the development of modern society. Alienation in this sense is characterized by the structure of possession and subsequent disappropriation of man's original constitution. Taking a closer look at Rousseau, it can be pointed out that he works with a simplified version of alienation in the form of hypothetical possession – disappropriation. In contrast, the general structure of alienation elaborated by Marx might be described with the possession – disappropriation – reappropriation formula. What we find in Rousseau can only be labeled "alienation" in a looser sense, since the semantic core of alienation, i.e. something's becoming strange or foreign to someone, does not exactly correspond to Rousseau's basic problem. He develops an idea that might be called "alienation", but has a significantly simpler conceptual structure than alienation in Marx and in the tradition relying on him.

As to conceptual clarifications, I have to remind that there is no explicit use of "alienation" in Rousseau's work. The only candidate for a conceptual antecedent of the term is the corrupted state of modern civilized humans, more precisely, modern man's distance to the original natural state. Rousseau's theory, thus, claims that modern civilized man had become alienated from man's original nature. Working with this rudimentary definition of "alienation", we have to work out the components corresponding to the semantic core of the term: it consists in something's becoming strange to someone what previously belonged to it. There is a clear restriction on the being that comes alienated, in as much

<sup>1</sup> Zehnpfennig 2013. 179; Jaeggi 2014; Struma 2001. 161. Christoph Henning thinks that key motifs of Rousseau's complex work could be arranged around the center of "the major topic of alienation" (Henning 2015. 35–36).

as it must have the cognitive capacity of recognizing strangeness and familiarity. Let us look at how Rousseau describes modern society and the state of modern, civilized man.

As mentioned above, Rousseau's two early discourses expound perhaps the most radical critique of civilization. He even goes beyond Plato's notorious critique of poetry which is far more restricted in its scope, since Rousseau holds the entire field of sciences, arts, and even morals to be corruptive. The first of the two *Discourses* from 1750 and 1755 works out principal objections against sciences, fine arts, and morals in general, whereas the *Second Discourse* tries to show how the development of human society creates fatally distorting conditions.

The *First Discourse* seeks not only to show that sciences and fine arts are luxurious and superfluous, but even that they might be regarded as morally perverting. Sciences and artistic production are luxurious, in so far as they presuppose free time and the suspense of efforts to survive. Furthermore, sciences and fine arts are perverting, since they do not only amount to wasting time, precious time that could be spent instead with other prestigious activities, but they make people more dependent on one another and make them seek recognition.<sup>2</sup> Consequently, Rousseau argues, cultivation of sciences and fine arts inevitably leads to the weakening of morals and human character. This principal objection to sciences, arts, and morals claims at the same time that they contribute to the maintenance of socially constructed false appearances. For Rousseau, these appearances are total and ubiquitous. In light of this, it is somewhat surprising that there are extraordinary personalities, e.g. Socrates, who are able to neutralize the negative effect of society upon them, and thus to step out of it. It is not made clear how this self-liberation from socially produced appearances in the case of Socrates is possible, and so it remains disturbingly vague how the individual's resistance to society's negative influences is possible.

It is the predicament of modern man to live in appearances which reproduce day by day his situation of mutual slavery. In important passage of the *First Discourse* Rousseau writes:

While the Government and the Laws see to the safety and the well-being of men assembled, the Sciences, Letters, and Arts, less despotic and perhaps more powerful, spread garlands of flowers over the iron chains with which they are laden, throttle in them the sentiment of that original freedom for which they seemed to born, make them love their slavery, and fashion them into what is called civilized Peoples. Need raised up Thrones; the Sciences and Arts have made them strong. (Rousseau 1997. 6.)

<sup>2</sup> As Günter Figal remarks, there is a contradiction in refusing the utility of science, on the one hand, and the project of a scientific contribution to the least developed knowledge, to that of self-knowledge, on the other hand (Figal 1991. 101).

One might discern here two basic propositions: there is (1) original freedom which (2) is transformed into unnoticed slavery in a way that the result is appealing to civilized man. To claim this, Rousseau needs to demonstrate original freedom, and the hypothesis of the original state serves exactly to justify that freedom. Rousseau makes here a fundamental assumption: Existing as a free being before, man becomes, in fact, a slave in modern civilization. In a perplexing manner, this is a slavery which remains unnoticed. This fact is underlined by the idea that pleasant intellectual achievements repress the feeling of freedom.

In a footnote, Rousseau gives a clue to a better understanding of the aforementioned slavery: there is a wide range of acquired needs the satisfaction of which makes us dependent on others. The longing for superfluous things, i.e. the acquired needs, as opposed to natural ones, are a “chaining” of man that is made clear by a contrast: “what yoke could be imposed upon men who need nothing?” (Rousseau 1997. 6). It should also be noticed that this idea makes sense obviously only under the assumption that we have a clear conception of basic and natural needs. Furthermore, for Rousseau superfluous, acquired needs come from the process of civilization. To describe these basic and natural needs, he uses therefore the hypothesis of the original state of savage man developed only later in the *Second Discourse*.

The ideological character of sciences and fine arts suggested by the passage is, however, not at all understandable. If they are pleasant, then they must be pleasant in themselves. But in this case it remains unclear how they could sweeten slavery, since Rousseau seems to suggest a kind of exchange. The characterization of modern man as a slave is, for sure, an overstatement which basically could not be compensated by such pleasures. Rousseau simply neglects the new possibilities of action like sciences and fine arts enabled by the division of labour, and he overemphasizes instead the mutual dependence implied by it. Let us see in more detail whether the description of the development of human society in the *Second Discourse* gives a better understanding of what could be called alienation in our context.

In contrast to the *First Discourse*, the *Second Discourse* gives an account of the point of reference on the basis of which modern society is evaluated. Furthermore, Rousseau develops here the point of inequality which turns out to be the basis of the critical assessment of sciences and fine arts. In the Preface to the opera *Narcisse* he underlines that inequality is both precondition (inequality of leisure) and major goal (prestige and distinction) of scientific and artistic activity.<sup>3</sup> It has to be added that the condemnation of scientific and artistic activity

<sup>3</sup> “A taste for letters always heralds the beginning of corruption on a people [...] For, in an entire nation, this taste can only rise from two sources, both of them bad, and both of them perpetuated and increased by study, namely idleness and a craving for distinction” (Rousseau 1997. 97). It is worth mentioning that Rousseau’s view is similar to that of Freud’s sublimation thesis in tracing back scientific and artistic activity to motivations that the actors would not



builds on the strong presupposition that the only goal or at least the primary goal of these activities is prestige, i.e. to distinguish oneself from others (Fetscher 1975. 20). We have, then, two major steps in Rousseau's description: first, he gives a picture of the savage man, and secondly, outlines a complicated process of the constitution of modern society.

The picture of the savage man serves to comprehend the original situation of man that had been abandoned step by step by the process of civilization. Rousseau's argumentation becomes at this point ambiguous. He is undoubtedly aware of the special difficulties implied in reaching a picture of natural human beings.<sup>4</sup> In the Preface to the *Second Discourse*, he asks:

how will man ever succeed in seeing himself as Nature formed him, through all the changes which the succession of times and of things must have wrought in his original constitution, and to disentangle what he owes to his own stock from what circumstances and his progress have added to or changed in his primitive state? like the statue of Glaucus which time, sea, and storms had so disfigured that it less resembled a God than a ferocious Beast. (Rousseau 1997. 124.)

In trying to grasp the "original constitution" of human beings, Rousseau thinks necessarily to proceed on the assumption that men were naturally equal among themselves, and so the first origin of the differences between them needs to be found (ibid.). His solution of the problem consists of elaborating a sort of ideal measurement in the framework of a thought experience: "For it is no light undertaking to disentangle what is original from what is artificial in man's present Nature, and to know accurately a state which no longer exists, and about which perhaps never did exist, which probably never will exist, and about which it is nevertheless necessary to have exact Notions in order accurately to judge of our present state" (Rousseau 1997. 125). The description of the savage man, then, proves to be a theoretical device in order to be able to distinguish the natural from the artificial.

In spite of the hypothetical character of the original constitution, Rousseau believes to have found certain traces of the original human nature in earlier forms of culture. He tries to explain the deforming character of modern developed society also from these findings. Rousseau identifies already in the *Second Discourse* invariant determinations of human nature, first of all, perfectibility, love of the self (*l'amour de soi*) and compassion (*pitié*). He considers these elements to be part of human nature everywhere and every time, although they

---

admit and would not be able to recognize (craving for distinction in Rousseau, sexual satisfaction in Freud).

<sup>4</sup> There are some who do not reflect on the status of the original condition of humans, but simply take it as unproblematic.

might be weakened as can be seen in self-interest (*l'amour propre*). Universal features of human nature, however, cannot be observed immediately, they show themselves in human reactions and conduct. Rousseau's problem is that persons are not able to justify these reactions, and similarly, they can identify false needs without being able to demonstrate their falseness (Struma 2001. 73–74).

It is not necessary to follow the complicated declining process of the evolution of human society in order to focus on the essential point. It is that in Rousseau's description of the original, natural position we find an image of man which is hard to relate to what we think to be human life. The savage man has no moral qualities, no reflection and consequently only a reduced form of freedom, almost no sense of time, no self and self-awareness, and a peripheral attitude towards others (empathy put aside which cannot really be explained).<sup>5</sup> We have here a set of properties and abilities that on the one hand serve as a basis of critique of civilized human life, but on the other hand are not able to offer a plausible concept of human life in general.<sup>6</sup> The savage and the civilized man differ in their inmost intentions, inclinations, and desires:

The first breathes nothing but repose and freedom, he wants only to live and to remain idle, and even the Stoic's *ataraxia* does not approximate his profound indifference to everything else. By contrast, the Citizen, forever active, sweats, scurries, constantly agonizes in search of ever more strenuous occupations: he works to the death, even rushes toward it in order to be in a position to live, or renounces life in order to acquire immortality. (Rousseau 1997. 186–187.)

It is entirely unclear what the point of freedom could be in “living and remaining idle”. The problematic character of modern man appears, so Rousseau thinks, both on the level of the human species and of the individual. The latter aspect concerns the problem of authenticity which can be grasped but indirectly. Although we do not entirely know who we are, we are still able to realize when certain actions would be against our true selves – at least this is what Rousseau aims to establish.

Rousseau's proposal obviously has several weak points, since it cannot establish a distinction between two classes of feelings that could be traced back to the distinction of self-sufficient love of the self of the savage man and dependent,

<sup>5</sup> Barbara Zehnpfennig observes that empathy is simply against the logic of the natural state, since human beings live isolated and their contact with others is marginal (Zehnpfennig 2013. 180).

<sup>6</sup> Rousseau's picture of the savage man which he thinks to be anti-Hobbesian is far not so different as compared to Hobbes. Rousseau claimed in the *Second Discourse* that theories of an original contract in a situation before any society made the mistake of projecting modern man distorted by society into a position before society. See on this point Wolfgang Kersting's comments on *Contrat social* (Kersting 2002. 20).

egoistic *amour propre* of modern man. More importantly, a second weak point is the general refusal of any kind of comparison and competition in human life which seems to exaggerate their disadvantages. Rousseau makes comparison responsible for the mutual dependence of human beings understood as slavery; what he establishes, however, is the fact that in stating or articulating a need or a feature we need others since we rely on their agreement as a kind of guarantee. It is mixing up two functions if intersubjective reliability becomes identified with the relationship of mutual dependence.

Furthermore, society for Rousseau is something that cannot have but destroying effects. This view is obviously reductive, since division of labor is presupposed in various higher intellectual achievements that could not be accomplished without satisfying biological needs with the help of others. The narrow-mindedly negative estimation of these achievements is the prize Rousseau is apparently ready to pay in order to have a perspective to criticize comfort, luxuriousness, and abundance. The ideal of frugality underlying Rousseau's critique enables him to refuse negative social tendencies in human history, but it cannot, in turn, allow higher intellectual achievements. The unreality and implausibility of the original state of humans build the major difficulty in talking about alienation in Rousseau since it simply makes for conceptual reasons impossible to overcome the alienated situation in the sense of returning or reconstructing it. Non-alienated human life would be not human at all.<sup>7</sup>

Concluding this section, it has to be settled that a wider conception of alienation lies in Rousseau's harsh disdain of human culture. The semantic core of this alienation is the loss of original capacities and natural instinctiveness. What Rousseau did not show, except in a very hypothetical manner, is the identification of very human nature with original capacities and instinctiveness. The hypothetical character of the savage man does not even permit the question why culture and education cannot be part of the human essence.

The talk of alienation in Rousseau is made complicated by the fact, as indicated, that the point of departure of the process of becoming strange remains unspecified. If human nature is hypothetical, a thought experience as a methodological device, then alienation in a narrower sense cannot be said to have taken place. For this reason I propose to label Rousseau's description as a conception of alienation only in a broader sense. However, it does not mean a solution to the remaining problem – i.e. what is the status of human nature in Rousseau? It is open to debate whether Rousseau's critique of one-sided Enlightenment

<sup>7</sup> We have, in fact, two alienation-claims in Rousseau: first, that man gets alienated from nature, and second, that man gets alienated from his- or herself. From this angle, diremption (*Entzweiung*) is the basic problem of stepping out of the original state: diremption with itself and diremption with nature – two in one in Rousseau's version. On alienated fine arts see my paper (Olay 2017).

appreciation of rationality can be formulated in a less radical way so that it does not fall into a similarly one-sided overestimation of feeling and sentiment.

Let us turn to a brief sketch of Marx' theory of alienation.

## II. ALIENATION IN MARX

Rousseau described problematic features of modern man which could be labeled as alienation, but only in a broader sense. In contrast to his broader conception of alienation, a more specific theory can be found in the thought of Karl Marx. The following sketch does not aim to give an exhaustive account of alienation in Marx but enables us to see important conceptual differences between his theory and Rousseau's.

In Marx's work, we find a shift from alienation in the early Paris Manuscripts (*Economic and Philosophical Manuscripts*, *Ökonomisch-philosophische Manuskripte*) to reification / objectification in the later work (*Kritik der politischen Ökonomie*). It is debated whether this means a break in treating the issue, or even abandoning it, or rather implies the presence of the topic in the whole work.<sup>8</sup> Be it as it may, normative basis of alienation for the young Marx is the concept of man's self-realization in the working process: "labor is the self-realizing human activity". The self-realization, in turn, takes place in a double movement of a prior objectification and a following re-appropriation.

Famously, the early Marx claimed that labour in capitalism cannot be but alienated. He talks about alienation of the worker in four different sense: he is alienated a) from the product of his work, b) from the process of his working, c) from species-being (*Gattungswesen*) – i.e. man is not exercising activities proper to true human nature and capacities –, and finally d) from others. Considering the inner dependence of these forms, the essential point can be found in the second one, since the first alienation is a consequence of the "alienation within the activity of work itself" (*Entfremdung... in der Tätigkeit der Arbeit selbst*) which is a kind of self-alienation (*Selbstentfremdung*, 515) of the worker. Talking about

<sup>8</sup> On various positions see Kolakowski 1978, 263ff. It lies beyond the scope of this paper to clarify the reasons for the shift from the early manuscripts' description of alienation to the later works. Kolakowski makes a case for the continuity thesis by claiming that the Paris Manuscripts "are in effect the first draft of the book that Marx went on writing all his life, and of which *Capital* is the final version" (Kolakowski 1978, 132–33). Tilman Reitz's proposal interestingly differentiates between what philosophy definitely cannot offer (revolutionary changes) and its actual functions (to support ideological agreement with the existing order). Kübler remarks that we do not find any justification of the refusal of capitalism in the later work, only in the *Manuscripts* (Kübler 2013). Zehnpfennig claims that there is no strict separation of the alienation-theorem and the later critique of capitalism: "Seine im *Kapital* entwickelte Kapitalismuskritik und seine Revolutionstheorie lassen sich im Grunde gar nicht verstehen, wenn es nicht die in der Entfremdungstheorie beschriebenen Defizite wären, die durch die Revolution behoben werden sollen." (Zehnpfennig 2013, 185).

self-alienation means that the working activity is “independent” (*äußerlich*) from the worker, it does not belong to his essence, it is forced labour (*körperlich und geistig ruinöse Zwangsarbeit*), so that it is the exact opposite of work as self-realization in the sense of “free psychic and intellectual energy”.<sup>9</sup> We skip the question whether everything we call labour or work must have these features or not.<sup>10</sup>

Marx’s conception of work as the opposite of self-realization contains the characterization of work as “abstract”. He follows here Adam Smith’s description of the poverty of workers, and considers his identification of work with pain as naturalization of alienated work. Marx regarded property as something that should be explained, not simply accepted, as leading figures of political economy like Smith and Locke did. Whereas he explicitly acknowledges categories and “laws” of national economy, he refuses it as being an ahistorical perspective without offering a basic principle for the explanation of property.<sup>11</sup>

Without entering further into the complexities of Marx’s conception of alienation, it can be stated that he thinks the transformation of alienated work into a non-alienated situation possible.<sup>12</sup> Provided that alienated work can be traced back to private property, it is consequent to see the main purpose of the process of history, in a situation without private property, i.e. in Communism. In the present context, it is enough to emphasize that even if the realization of Communism might be regarded as problematic from a practical point of view, it cannot be doubted that in Marx’s eyes it would mean a non-alienated state. With this we have a basically different semantics of alienation in Marx as compared to Rousseau, since Marx thinks a sequence of possession – disappropriation – re-appropriation possible, whereas the latter has but a short version in the form of hypothetical possession – disappropriation.

Let us turn now to Lukács’s theory of alienation and reification.

<sup>9</sup> “Der Arbeiter fühlt sich “nicht wohl, sondern unglücklich [...], [...] fühlt sich daher erst außer der Arbeit bei sich und bei der Arbeit außer sich” (514) (quoted in Elbe, 6).

<sup>10</sup> It is not here to discuss an alternative conception to this in the work of Mihály Csíkszentmihályi. His deeply Aristotelean conception of “flow” develops the basic point that each activity, even monotone and mechanic ones, might be the source of a pleasant contentedness.

<sup>11</sup> “Die Nationalökonomie geht vom Faktum des Privateigentums aus. Sie erklärt uns das-selbe nicht.” (510) Sie fixiert “die *entfremdete* Form des geselligen Verkehrs als die *wesentliche* und *ursprüngliche* und der menschlichen Bestimmung entsprechende” (451).

<sup>12</sup> Kolakowski regards a series of “critiques” of Marx – including among others the *Paris Manuscripts* and *Capital* itself – as more and more elaborated versions of the same basic idea which he formulates as follows: “We live in an age in which dehumanization of man, that is to say the alienation between him and his own works, is growing to a climax which must end in a revolutionary upheaval; this will originate from the particular interest of the class which has suffered the most from dehumanization, but its effect will be to restore humanity to all mankind” (Kolakowski 1978. 262).

### III. LUKÁCS'S THEORY OF ALIENATION

Lukács's contribution to the theory of alienation has also often been seen in his concept of reification (*Verdinglichung*) – the heading of his critique of capitalist society. As indicated at the outset, I will focus in this section on the clarification of how non-reified or de-reified conditions are, rather implicitly, described by Lukács. The presupposition of a concept of non-reified or non-alienated conditions lies at the heart of every theory of reification or alienation including Marx' conception, too. By the discussion of the concept of reification developed in the chapter "The Phenomenon of Reification" in *History and Class Consciousness* I try to show that Lukács's contribution to the theory of reification lies not in a proposed solution, but rather in differentiation and extension of the phenomenon or reification along broader social dimensions. At the same time, Lukács, as Marx before him, still owes an answer to the question how non-reified relations and non-alienated conditions should be conceived of.

As to Lukács's analysis of reification, his famous conception in *History and Class Consciousness* has proved to be one of his most influential ideas. His contribution to the theory of alienation has also often been seen in his concept of reification (*Verdinglichung*). With his concept of reification Lukács not only "found out", as it were, what came to be published in Marx's *Economic and Philosophic Manuscripts* only nine years later, but continued at the same time to develop his "romantic anti-capitalism" from his pre-Marxist period.

The first important point with regard to *History and Class-Consciousness* is that Lukács declares to revive Marx's method in Hegelian spirit. As explained in *What is Orthodox Marxism?*, Lukács sees the center of Marx's thought in the demand for a revolutionary transformation of the world. With this move against the main line of the Second International, the core of Marxism is grasped as an activist, revolutionary attitude towards the existing conditions, instead of the scientific-economic self-interpretation of the late Marx. Lukács touches here a sensible point in the Marxist tradition, viz. the tension between economic analysis of capitalism and class-struggle in Marx's conception. The ambiguity of an activist-voluntarist strand and an economic-scientific strand could be traced back to the early writings of Marx. Lukács himself, however, does not hesitate to make the fundamental presupposition that late capitalist society needs revolution, not only political ameliorations and amendments. It is not easy to isolate for what reasons he entertains this conviction. Lukács possibly takes it over from Marx himself who was persuaded of the inevitability of revolution, too.<sup>13</sup> For Marx, the idea depends on the structural problems of capitalist production he considers to be irreparable by a step-by-step procedure or evolution.

<sup>13</sup> Thesis 11 on Feuerbach.

Lukács's collection of essays is basically a reaction to the theoretical crisis of Marxism after World War I. The crisis comes from the fact that the proletariat, against Marx's predictions, does not seem to bring revolutionary changes, and seems even less to move towards a revolution. Still worse, social democracy appears as an alternative reaction, both theoretical and practical, to the fact that revolution does not arrive. *History and Class Consciousness* is, thus, to a high extent a political work, and some features of Lukács's Marxism are consequences of this. First of all, the significance of dialectics as primacy of the whole against the parts needs to be underlined. As Lukács puts it, "[t]his absolute primacy of the whole, its unity over and above the abstract isolation of its parts – such is the essence of Marx's conception of society and of the dialectical method" (Lukács 1971. 27). In terms of this reading of dialectics, he takes the Marxist method as the attempt to consider the social world as a single whole of "totality".<sup>14</sup> In doing so, the underlying premise is "the belief that in Marx's theory and method the true method by which to understand society and history has finally been discovered". For Lukács, then, the Marxist method serves the pre-eminent aim of the "knowledge of the present" (Lukács 1971. xliii).

As a second essential moment, the explicitly revolutionary aspect of Lukács's reading of Marxian dialectics should also be accentuated. To understand society and history, the "knowledge of the present" is not merely theoretical and contemplative, as clearly indicated by Marx's eleventh thesis on Feuerbach which is the motto for the study on orthodox Marxism: "The philosophers have only interpreted the world in various ways; the point, however, is to change it" (Lukács 1971. 1). Correspondingly, a revolutionary action is prepared by "a dialectical knowledge of reality, which discovers the tendencies pointing towards the ultimate objective not in isolated facts, but in the dynamic totality" (Löwy 1979. 174). It is within this theoretical framework that the central essay – "Reification and the consciousness of the proletariat" – should be understood: the two major components of the title express well the theoretical program. "Reification" stands for the description of the crisis of capitalist society, and "the consciousness of the proletariat" is the revolutionary impetus which needs to be actualized in order to overcome reification.

Lukács's concept of reification (*Verdinglichung*) is a theory of objectified or reified relationships that relies on Marx's theory of commodity fetishism. Some have already suggested that reification is a special case of alienation, as a widespread form characteristic of modern capitalist society. The most important argument against this identification is the fundamentally different scope of al-

<sup>14</sup> "His view that this is the key to Marxist theory did not alter from 1919 to 1971. [...] Marxism, according to Lukács, would be impossible if it did not involve the principle that the social 'totality' cannot be reconstructed by accumulating facts. Facts do not interpret themselves: their meaning is only revealed in relation to the whole, which must be known in advance and is thus logically prior to the facts" (Kolakowski 1978 III. 265).



iation and reification. Alienation is apparently a much wider phenomenon than reification, since there are cases of alienation not being necessarily cases of reification, e.g. alienation from other human beings.

Lukács begins the explanation of reification with an analysis of commodity-structure which he states to be *the* basic problem of capitalist society. With a surprising universality, he declares that in the age of capitalist society “there is no problem that does not ultimately lead back to that question and there is no solution that could not be found in the solution to the riddle of commodity-*structure*” (Lukács 1971. 83). Lukács not only stresses the central character of commodity-structure but assumes its model-character for all aspects of capitalist society. The commodity-structure is the central, structural problem of capitalism because it yields a “model” of objective and corresponding subjective forms in bourgeois society (*ibid.*). The description of reification is, thus, grounded on the commodity-fetishism described by Marx in *Capital*. What complicates matters is that Lukács’s argumentation exhibits deep affinity also with Marx’s early theory of alienation, even if he could not know it. The question must be suspended here whether the perspective of *Capital* carries on the early writings on alienation as some think.<sup>15</sup> Nevertheless, in our context, however, it is worth noting that the theory of alienation in the young Marx made essential assumptions concerning a “human being”, whereas the theory of commodity fetishism doesn’t need such assumptions. Furthermore, the core of Marx’s idea of alienation is not an objectifying relationship that would make an object out of human skills, properties or human beings. The point of reification in Lukács’s sense is exactly this move of making something/somebody into an object or considering something/somebody as a mere object.

Even more important is the extension of the analysis of reification as compared to Marx. In Lukács’s view it is not only market and exchange processes, but all dimensions of capitalist society that show reification processes. In other words, he broadens the scope of the reification structure processes in capitalism that are, he adds, infinite in tendency. By extending reification to all aspects of society, he arrives at an overall diagnosis of his time. With regard to the phenomenon of alienation, the novelty in Lukács’s description of reification lies less in

<sup>15</sup> See for example Karl Korsch’s claim that what Marx baptized “self-alienation” in his early philosophical period, became “commodity fetishism” in his later critical-scientific period. See also Leszek Kolakowski’s comment: “Although the word ‘alienation’ occurs less often, the theory is present in Marx’s social philosophy until the end of his life; ‘commodity fetishism’ in *Capital* is nothing but a particularization of it. When Marx writes that commodities produced for the market take on an independent form, that social relations in the commercial process appear to the participants as relations among things over which they have no control (exchange value being falsely represented as inherent in the object and not as an embodiment of labour), and that the supreme type of this fetishism is money as a standard of value and means of exchange – in all this Marx is reproducing the theory of self-alienation that he had formulated in 1844.” (Kolakowski 1978. I. 173.)



new forms or variations, but in the universality of reification in all social forms and dimensions of capitalist society.

The core of the phenomenon of reification is that a relation between human beings “takes on the character of a thing and thus acquires a ‘phantom objectivity’, an autonomy that seems so strictly rational and all-embracing as to conceal every trace of its fundamental nature” (Lukács 1971. 83). As already indicated, Lukács does not confine his analysis to the economic sphere, but tries to show that it is necessary “for the commodity structure to penetrate society in all its aspects and to remould it in its own image” (ibid. 85). Although he seems to promise here a kind of justification for this penetration, there is no real explanation, not even an attempt to spell out why the thing-structure should become pervasive in every dimension of capitalist society. The lack of explicit explanation is particularly unfortunate since the connection of the economic sphere with other dimensions of society, the one-sided dependence of the latter on the former was an often criticized idea in Marx’s oversimplifying base-superstructure scheme. One might object that there are fields – e.g. human relationships such as friendship – that are, or at least, can be resistant to commercialization and commodification.

Lukács’s comment on the famous Marxian passage on the fetishism of commodity helps to highlight his position: “a man’s own activity, his own labor becomes something objective and independent of him, something that controls him by virtue of an autonomy alien to man” (ibid. 87). The argumentation, then, differentiates between an objective and a subjective side of the phenomenon. In terms of his example of the unchangeable, but knowable laws of market, Lukács suggests that his very problem is not the strange character of reified phenomena, but the independence of reified phenomena and man’s loss of influence upon them. In contrast to this, as we saw above, alienation in Marx is a kind of distanciation from different aspects of the working activity, but not an objectifying relationship that would make an object out of human factors or human beings.

The specific negative evaluation of this objectifying relationship is not really justified by Lukács. The single fact that we regard human capacities, performances as properties of objects could not yet warrant a negative evaluation. Axel Honneth also stresses that the type of reification is unclear, since Lukács misses to specify whether it is an epistemic category mistake, morally wrong behavior, or a distorted form of praxis (Honneth 2008. 25–27). Lukács’s point on the negativity of reification is that the worker loses its organic relationship to his or her own skills and capacities: “With the modern ‘psychological’ analysis of the work-process (in Taylorism) this rational mechanisation extends right into the worker’s ‘soul’: even his psychological attributes are separated from his total personality and placed into specialised rational systems and their reduction to statistically viable concepts” (Lukács 1971. 88). He thinks the rational fragmentation of “the subjects of labour” to be far-reaching, both individually and

collectively. The objectification of the worker's labour-power into something opposed to his total personality becomes now a permanent reality of his daily life. And Lukács adds:

Here, too, the personality can do no more than look helplessly while its own existence is reduced to an isolated particle and fed into an alien system. On the other hand, the mechanical disintegration of the process of production into its components also destroys those bonds that had bound individuals to a community in the days when production was still "organic". (Ibid. 90.)

What makes Lukács's analysis distinctively different from the commodity-fetishism, is an additional essential aspect which had been inspired by Max Weber. Weber connected the process of rationalization with specialization, and this connection is especially important for Lukács: "the principle of rationalisation based on what is and *can be calculated*" (ibid. 88). Rationalization, in his view, intensifies the process of reification:

...the principle of rational mechanisation and calculability must embrace every aspect of life. Consumer articles no longer appear as the products of an organic process within a community [...] They now appear, on the one hand, as abstract members of a species identical by definition with its other members and, on the other hand, as isolated objects the possession of which depends on rational calculations. Only when the whole life of society is thus fragmented into the isolated acts of commodity exchange can the "free" worker come into being. (Ibid. 91.)

It is interesting to note that Lukács doesn't really explain the necessity of rationalization in the production process; he simply claims it, and goes on to an argument we already find in Marx about the anarchic nature of capitalism, viz. that capitalist production seeks profit and doesn't follow real needs of a real community.

While integrating Marx and Weber, Lukács claims that commodity production revolutionizes the production process. He combines here two traditions, in so far as he adds to the Marxian critique of capitalism the dimension of philosophy of life in the form of a rather unorthodox reading of Weber's rationalization thesis. This combination is even stranger, the more clearly we see that Weber attempted an explanation of capitalism in contrast to Marx. It is, however, less clear how the two threads of argumentation intensify each other. To put it otherwise, it is undecided which explanatory factors stem from Marx and which from Weber.

The central claim of Lukács is, then, that in capitalism reification becomes the second nature of man. He asserts that human beings in capitalism inevitably get accustomed to perceiving themselves and their environment as mere

objects. Lukács concentrates here on transformations on the subject's side, especially on transformations under the pressure of commodity exchange. Persons under conditions of permanent commodity exchange, he suggests, change their basic attitude to their whole environment, in so far as they acquire a contemplative stance, they become "detached observers" of their own existences which are "reduced to an isolated particle and fed into an alien system" (ibid. 90). By contemplative attitude, Lukács means the aspect of the passivity of the observer who is contemplating the independent processes where he or she does not grasp himself or herself as an active participant of what happens.<sup>16</sup> Interestingly, Lukács considers the structure of detachment, viz. "the split between the worker's labour power and his personality" a pervasive feature of every field of capitalist society (ibid. 99).

With the claim that capitalist society has arrived into a final stage of reification, Lukács reproduces a similar diagnosis to that of Marx. The criteria to judge that society has entered into a final stage are *eo ipso* precarious, even if they carry a heavy burden of proof. In fact, the final, irreversible character of capitalist society is the reason why Lukács, as already mentioned at the outset, does not even consider the possibility of a step-by-step or piecemeal improvement of society. There is no other way out of this situation than a revolution of the proletariat, and Lukács's efforts are directed from this point on to solve theoretical difficulties with regard to this revolution.

Two main connected difficulties arise for him. First, the proletariat in its reified status should be revolutionized, and secondly, in order to solve the first problem, a non-reified point of departure is needed. Lukács presupposes that it is impossible to change society's reified status from within so that a factor not touched by reification is needed to initiate the process of dereification. For this purpose, he follows Lenin's proposal concerning the role of a political avant-garde embodied by the Communist Party. The Communist Party should be the non-reified beginning of the revolution disembarassing from society's reification. But this is a theoretical requirement, not a factual description. And it is, finally, the reason why Lukács's description of the Communist Party is entirely unreal, unfounded Romanticism. In our context, however, we cannot follow his theory of the party in detail. Let us turn to the conclusion.

<sup>16</sup> See Honneth's remarks: "Unlike Martha Nussbaum, Lukács isn't interested in determining the point at which the reification of other persons becomes a morally reproachable act. Instead, he sees all members of capitalist society as being socialized in the same manner into a reifying system of behavior, so that the instrumental treatment of others initially represents a mere social fact and not a moral wrong" (Honneth 2008. 26).

#### IV. CONCLUSION

The overview of these three conceptions served to make clear the basic structure of alienation, following the question of what should be reappropriated in overcoming alienation and how we experience our being alienated. It has been shown that Rousseau developed a simplified version of alienation in the form of hypothetical possession – disappropriation, whereas Marx elaborated the general structure of alienation that might be described with the possession – disappropriation – reappropriation formula. Furthermore, it has been argued that the novelty of Lukács's analysis in contrast to Marx lies in the extension of the scope of reification, since he thinks that not only market and exchange processes, but all dimensions of capitalist society show reification processes which are in capitalism infinite in tendency.

Two conclusions should be stressed here: the problem of non-alienated conditions, on the one hand, and the neglect of the individual's individuality, on the other hand. It is easy to see that the problem of describing non-alienated conditions remains a hard theoretical nut to crack, as can be seen from contemporary examples. Some contemporary thinkers, mainly in the tradition of the Frankfurt School, pay special attention to alienation. Axel Honneth, Rahel Jaeggi, Hartmut Rosa, János Weiss published in the last decade books on alienation. The usage of alienation and reification is somewhat confuse, or at least, complex. For example, it is characteristic that Jaeggi suggests relying on Heidegger to distinguish two aspects of self-alienation: it means, first, to make oneself to a thing, and second to adjust one's decisions and conduct to what others do (Jaeggi 2005. 38) In doing so, Jaeggi mixes alienation with authenticity and tries to integrate the latter problem into the former. Her case leads to the second conclusion.

Descriptions of alienation and reification processes are relevant and interesting for an existential analysis of human beings. They argue against a – conscious or unconscious – reduction or objectification of distinctively human features. In doing so, they theorize and defend what is human in human beings. However, they characteristically lack a sensibility for the individuality of the individual. It seems to be a consequence of the focus on alienation and reification that the distinctive particularity of the human individual cannot be grasped sufficiently. The descriptive interest in what is essentially human loses sight of individuality. Therefore the theories of alienation remain to be complemented with an account of what it is like to be an individual.

## REFERENCES

- Elbe, Ingo 2014. Entfremdete und abstrakte Arbeit. Entfremdete und abstrakte Arbeit. Marx' *Ökonomisch-philosophische Manuskripte* im Vergleich zu seiner späteren Kritik der politischen Ökonomie. In Ingo Elbe – Philip Hogh – Christine Zunke (eds.) *Oldenburger Jahrbuch für Philosophie 2012*. Oldenburg, BIS-Verlag, 7–69.  
[http://www.rote-ruhr-uni.com/cms/IMG/pdf/Entfremdete\\_Arbeit.pdf](http://www.rote-ruhr-uni.com/cms/IMG/pdf/Entfremdete_Arbeit.pdf); last access 16. 08. 2016.
- Fetscher, Iring 1975. *Rousseaus politische Philosophie*. Frankfurt/M, Suhrkamp.
- Figal, Günter 1991. Philosophische Zeitkritik im Selbstverständnis der Modernität. Rousseaus *Erster Discours* und Nietzsches zweite *Unzeitgemäße Betrachtung*.
- Henning, Christoph 2015. *Theorien der Entfremdung*. Hamburg, Junius.
- Honneth, Axel 2008. *Reification. A New Look at an Old Idea*. Oxford, Oxford University Press.
- Jaeggi, Rahel 2014. *Alienation*. New York, Columbia University Press.
- Iorio, Marco 2013. Veränderung, Verdinglichung, Entfremdung. Über Marxens verhege-It-verhagelte Ontologie. In Rahel Jaeggi – Daniel Loick (eds.) *Karl Marx – Perspektiven der Gesellschaftskritik*. Berlin, Akademie Verlag.
- Kersting, Wolfgang 2002. *Jean-Jacques Rousseaus "Gesellschaftsvertrag"*. Darmstadt, Wissenschaftliche Buchgesellschaft.
- Kolakowski, Leszek 1978. *Main Currents of Marxism*. Vol. 1–3. Oxford, Clarendon Press.
- Kübler, Lukas 2013: Marx' Theorie der Entfremdung; In Rahel Jaeggi – Daniel Loick (eds.) *Karl Marx – Perspektiven der Gesellschaftskritik*. Berlin, Akademie Verlag.
- Lohmann, Georg 1991. *Indifferenz und Gesellschaft. Eine kritische Auseinandersetzung mit Marx*. Frankfurt/M, Suhrkamp.
- Löwy, Michael 1979. *Georg Lukács – From Romanticism to Bolshevism*. London, NLB.
- Lukács, Georg 1971. *History and Class Consciousness. Studies in Marxist Dialectics*. Cambridge/MA, MIT Press.
- Marks, Jonathan 2005. *Perfection and Disharmony in the Thought of Jean-Jacques Rousseau*. Cambridge, Cambridge University Press.
- Marx, Karl 1968. *Ökonomisch-philosophische Manuskripte aus dem Jahre 1844*. In *Karl Marx – Friedrich Engels Werke*. Vol. 40. Berlin, Dietz Verlag, 465–588.
- Olay, Csaba 2017. Alienation in Rousseau. In Gábor Boros – Judit Szalai – Olivér István Tóth (eds.) *The Concept of Affectivity in Early Modern Philosophy*. Eötvös Kiadó, Budapest, 271–283.
- Rousseau, Jean-Jacques 1997. *The Discourses and Other Early Political Writings*. Cambridge, Cambridge University Press.
- Struma, Dieter 2001. *Jean-Jacques Rousseau*. München, C. H. Beck.
- Zehnpfennig, Barbara 2013. Rousseau und Marx: das Ende der Entfremdung. In Oliver Hidalgo (ed.) *Die langen Schatten des Contrat social. Demokratie und Souveränität bei Rousseau*. Wiesbaden, VS Verlag.
- Zintl, Reinhard 2009. Rousseau. In Michael Becker – Johannes Schmidt – Reinhard Zintl (eds.) *Politische Philosophie*. Paderborn, Schöningh, 64–76.

PHILIPP HÖFELE

## New Technologies and the “Heuristics of Fear”: The Meaning and Prehistory of an Emotion in Jonas, Heidegger and Hegel\*

I. INTRODUCTION: THE “HEURISTICS OF FEAR” AS  
A METHODOLOGICAL FEELING FOR TECHNOLOGY ASSESSMENT

In the present age, a constant evaluation of technology cannot be put off. Thus, in view of the ever-increasing dominance of technology within the biosphere of our planet, one already speaks of a distinct “technosphere” (cf. Zalasiewicz et al. 2017). Not only is the development of new technologies accelerating at an ever faster pace, as the term “Great Acceleration” suggests (cf. Steffen et al. 2015), but the boundaries between the natural and the artificial are becoming increasingly blurred. This has led to the assumption of a new geological epoch, the “Anthropocene” (cf. Crutzen 2000). It is not least against this background that technology assessment has developed a variety of methods in recent decades to prevent the dangerous excesses of technological development at an early stage (cf. Grunwald 2002). However, since it is ultimately a matter of dealing with the ignorance of future consequences, no matter how precise and complex the methods of forecasting may be, there is still a gap that by definition cannot be filled by discursive knowledge. From the very beginning of technology assessment, emotional knowledge has always been used alongside rational, mathematically and statistically determinable forms of knowledge.

The best known approach is surely Jonas’ “heuristics of fear” (cf. von Sass 2016). Admittedly, this principle proposed by Jonas has often been dismissed as too far-reaching and even as a form of conservatism which, for the sake of preserving the existing, tries to exclude every conceivable risk (cf. Grunwald 2002. 214; Schmidt 2013. 146). However, this principle can by no means be dismissed as obsolete, since the increasing technical possibilities make such a radical precautionary principle seem more important than ever before (cf. Böhler 2008).

\* Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy – EXC-2193/1 – 390951807. I thank James Fisher very much for the correction of the English text.

Moreover, the implications of this Jonasian "heuristics of fear" have rarely been fully explored, and these become first and foremost visible when one considers the precursors of this feeling in Hegel and especially Heidegger. In both "forerunners" this feeling is neither exclusively nor primarily a characteristic of a negative-pessimistic world behaviour. On the contrary, both the idealist Hegel and the hermeneut Heidegger understand fear or anxiety as guaranteeing the possibility of liquefying the given, which should ensure openness for an actual future or even a new level of consciousness.

The historical interest in tracing the prehistory of the Jonasian concept of a "heuristics of fear" in Hegel and Heidegger, which will be pursued in the following, is thus also accompanied by a systematic concern: namely, to explore the significance of a theory of feelings for technology assessment. It is the thesis advocated here that the "affective element" (Jonas 1979. 165) in (technological) ethics emphasized by the "heuristics of fear" can only be adequately understood against this historical background. The "sense of responsibility" (*Verantwortungsgefühl*) (ibid.) articulated in this fear proves to be such a sense, which opens up a horizon of possibilities for responsibility and is not intended to negate possibilities due to an allegedly exaggerated sense of caution.

In the following, it is firstly necessary to deal with the concept and function of a "heuristics of fear" in Jonas' work *The Imperative of Responsibility* from 1979 (chap. 2). Afterwards, we will discuss the concept of *Angst* in Jonas' teacher Martin Heidegger. This clearly forms the model for Jonas' concept of fear, although Jonas, as the preference of the concept of fear (*Furcht*) over that of anxiety (*Angst*) already indicates in a purely external sense, is striving to distinguish himself from his teacher in decisive points (chap. 3). A decisive demarcation is that Jonas, in contrast to Heidegger, is concerned with a form of "selfless fear" (Jonas 1979. 392; cf. Jonas 1984. 162). Although religious connotations conveyed through Kierkegaard also play a role here, this alteration in Jonas' work indicates his proximity to a thinker, namely Hegel, from whom, because he is a precursor of Marxism and Bloch, Jonas tends to distance himself. Thus, in the chapter of his *Phenomenology of the Spirit* entitled "Lordship and Bondage" (*Herrschaft und Knechtschaft*), Hegel also speaks of the feeling of fear in the case of a technical-craft process, a feeling that negates one's own self in favor of a higher level of consciousness (chap. 4). Even if, in contrast to his reception of Heidegger, it cannot be clearly proven whether Hegel's text also served as a model for Jonas, its parallelism allows the implications of Jonas' conception of a "heuristics of fear" to be interpreted even more clearly and comprehensively, as we will show in a concluding chapter (chap. 5).

## II. JONAS' "HEURISTICS OF FEAR" AS PRINCIPLE AND METHOD OF HIS FUTURE ETHICS

At the center of Hans Jonas' *Search of an Ethics for the Technological Age* (*Versuch einer Ethik für die technologische Zivilisation*) – the subtitle of his 1979 book *The Imperative of Responsibility* (*Prinzip Verantwortung*) – is the feeling of fear, which in its negativity is attributed a fundamental methodological significance for the new ethics promised by the book. The foreword to the 1984 English edition of the book refers to this central position of the “heuristics of fear” by introducing the term immediately after that of responsibility as a correlate of the newly gained human power to act (cf. Jonas 1984. x). The significance of Jonas' “heuristics of fear” for his ethical “search” is shown not least of all in the opposition to Ernst Bloch's *Principle of Hope* (*Prinzip Hoffnung*), published in 1954, whose utopian approach Jonas seeks to refute in its entirety (cf. esp. Jonas 1984. 194–201; Jonas 1979. 348–387).

Unlike Arthur Schopenhauer, for example, who places the feeling of compassion (*Mitleid*) at the center of his ethics, Jonas is not pessimistic (cf. Jonas 1984. 49; Jonas 1979. 101). In view of the fact that “the promise (*Verheißung*) of modern technology has turned into a threat (*in Drohung umgeschlagen ist*), or it has become inextricably linked with it”, as the initial thesis of his book states (Jonas 1979. 7),<sup>1</sup> he is interested in naming a “compass” (*Kompaß*) by which “first of all the ethical principles become discoverable (*entdeckbar*)” (ibid.). Only the “anticipated danger itself” (*vorausgedachte Gefahr selber*) and thus a “heuristics of fear” can provide this yardstick (Jonas 1979. 7f.). This upgrading of the “heuristics of fear” to the yardstick and principle of his ethics corresponds to the significance of the object of Jonasian “future ethics”, which does not only concern the “human fate” (*Menschenlos*) and its “physical survival”, but the “human image” (*Menschenbild*) and the “integrity of [his] essence” itself (Jonas 1979. 8).

This already shows that Jonas, with his new ethics, is not only concerned with a technical-ethical extension of the existing ethics in the area of technical risk assessment, but rather sees the new technical possibilities and the “changed [...] *nature of human action*” associated with them as calling into question the

<sup>1</sup> Jonas becomes clearer and more concrete in the preface to the English edition: “Not counting the insanity of a sudden, suicidal atomic holocaust, which sane fear can avoid with relative ease, it is the slow, long-term, cumulative – the peaceful and constructive use of worldwide technological power, a use in which all of us collaborate as captive beneficiaries through rising production, consumption, and sheer population growth – that poses threats much harder to counter. The net total of these threats is the overtaking of nature, environmental and (perhaps) human as well. Thresholds may be reached in one direction or another, points of no return, where processes initiated by us will run away from us on their own momentum – and toward disaster.” (Jonas 1984. ix.) Here Jonas already anticipates the later formulated theory of the so-called “tipping points”, from which there is no turning back once they have been reached. Cf. Lenton et al. 2019.



foundations of traditional ethics themselves: Neither "the human condition, determined by the nature of man and the nature of things, was given once for all", nor is "the human good" still "readily determinable", and above all "the range of human action and therefore responsibility" is not "narrowly circumscribed" any more (Jonas 1984. 1; Jonas 1979. 15). It is precisely this liquefaction and dissolution of the boundaries of the object of ethics that, in Jonas' view, forces us to ask the question of ethical "Principles and Methods" anew in the second chapter of *The Imperative of Responsibility*.

According to Jonas, it is in particular Kant's Categorical Imperative which is no longer sufficient for ethical reflection on the spatial and, above all, temporal delimitation of the radius of human-technical action. This imperative is directed towards a present that excludes the future (like the past), and according to this imperative there is "no self-contradiction in the thought that humanity would once come to an end"; for the purely logical "rule of self-consistency", which characterizes Kant's Categorical Imperative, is thus fulfilled, since in Kant's view unborn generations do not fall under the commandment of the self-purpose of currently responsible subjects (Jonas 1984. 11; Jonas 1979. 35). In this respect, Jonas sees the necessity of integrating a time horizon that can ultimately only be justified metaphysically into the Categorical Imperative, which, like Kant, he reformulates as follows in a fourfold variant:

"Act so that the effects of your action are compatible with the permanence of genuine human life"; or expressed negatively: "Act so that the effects of your action are not destructive of the future possibility of such life"; or simply: "Do not compromise the conditions for an indefinite continuation of humanity on earth"; or, again turned positive: "In your present choices, include the future wholeness of Man among the objects of your will." (Jonas 1984. 11; Jonas 1979. 36.)

In contrast to Kant's position, however, the possibility of these four imperative formulations cannot simply be derived from that of freedom as autonomy of the will (cf. Kant, AA 4, 446–455). Thus Jonas assigns to the "heuristics of fear" not only the task of naming the means of action required by the new ethics, but also the task of providing the metaphysical basis for the new imperative in its four variants. Even though Jonas refers again and again to the theological motive of reverence (*Ehrfurcht*), he does not want this foundation to be supported by reference to theological assumptions (cf. Jonas 1979. 8; 392f. Cf. also Huber 2018).

The "heuristics of fear" as "knowledge of the real and the probable in the realm of facts" is first introduced as a mediating sphere "between the ideal knowledge of ethical principles and the practical knowledge of political application" (Jonas 1984. 26; Jonas 1979. 62). The "heuristics of fear" is intended to mediate between the abstract knowledge of principles and their concrete application. At the same time, however, Jonas stresses that fear "is, rather, *heuristic*-

ly already needed within that doctrine [of the ethical principles] itself” (Jonas 1984. 26; Jonas 1979. 63). Analogous to the concept of *Angst* for Heidegger, as we will show in a moment, fear also reaches into the area of fundamental questions for Jonas, because it is precisely through the negative or consciously absent that a positive can be asserted. The possibility of the non-existence of something, according to the basic insight shared by Heidegger and Jonas, draws attention to the very existence or essence of something that in its self-evident presence is usually not conspicuous or remains unthematic. If we apply this to the possible dangers posed by technical progress, this means that it is precisely in imagining the destructive potential that could accompany this progress that we become aware of what is essential and worth preserving: “As long as the danger is unknown, we do not know what to preserve and why” (Jonas 1984. 27; Jonas 1979. 63). Since what is to be preserved is usually taken for granted, it first becomes noticeable when it no longer exists.

However, since the dangerous potential of technology is to be prevented, it is necessary to imagine this dreaded non-existence of something that needs to be preserved. And so the “‘First Duty’ of an Ethics of the Future” is just the “anticipatory conjuring up of this imagination” (Jonas 1984. 27; Jonas 1979. 64). In this context, “a casuistry of the imagination” is to be applied (Jonas 1984. 30; Jonas 1979. 67), which is not based on already known cases, but rather on those imagined in science fiction literature, for example. However, since this idea of a danger that could affect future generations has no potential for identification and therefore does not in itself cause fear, the second duty is the “bringing ourselves to this emotional readiness, developing an attitude open to the stirrings of fear in the face of merely conjectural and distant forecasts concerning man’s destiny” (Jonas 1984. 28; Jonas 1979. 65). Jonas thus demands a form of fear that is not at all self-evident, or is even paradoxical. For it is not a matter of “fear or anxiety for oneself” (*Furcht oder Angst um sich selbst*), but rather of “selfless fear” (Jonas 1979. 392; cf. Jonas 1984. 162), since this is directed toward a future humanity, but not toward one’s own presently living person. Only in this way can man do justice to his historical responsibility, “the flourishing of man in unconcerned humanity” (*das Gedeihen des Menschen in unverkümmelter Menschlichkeit*) (Jonas 1979. 393).

At first, it might seem as if Jonas is arguing here in an essentialist manner with regard to likeness to God established once and for all as the essence of the human being. Thus, at least in the original German edition, Jonas speaks conclusively of a “reverence (*Ehrfurcht*) for what man was and is, out of a shuddering retreat (*Zurückschauern*) at what he might become and which stares at us as this possibility from the imagined future (*als diese Möglichkeit aus der vorgedachten Zukunft anstarrt*)” (Jonas 1979. 393). That, however, a static-essentialist conception of man is not what Jonas is looking for is already shown by the fact that he deleted this theologically tinged final section in the English edition presented five

years later and only speaks of preserving the "integrity of his [man's] essence, which implies that of his natural environment" (Jonas 1984. 202). The fact that technology in modernity essentially defines human action and thus humanity, means for Jonas that with modern technology and its dangers, humanity is also under debate. Nevertheless, the "heuristics of fear" should not be accompanied by a conservative insistence on a supposedly timeless nature of human beings. Jonas already contradicts this insofar as he seeks to enrich Kant's present-fixed Categorical Imperative with a temporal component towards a future. It is precisely this fear that is intended to ensure this temporal reference of his ethics, which is oriented towards the preservation of humanity.

Jonas merely hints at how both can be thought of together. This only becomes fully understandable when one considers the background of this conception, namely the *Angst* conception of his teacher Martin Heidegger in *Being and Time*.<sup>2</sup> Even if Jonas at the same time resolutely dissociates himself from Heidegger's conception, especially from the self-fixation of Heidegger's *Dasein* (cf. esp. Jonas 1984. 88; Jonas 1979. 167), the principle-theoretical reevaluation of fear in Jonas unmistakably points back to Heidegger's analysis of *Angst*, as we will now show.

### III. ANGST AS A REVELATION OF THE POSSIBILITY HORIZON OF *DASEIN*

Heidegger's analysis of *Angst* in *Being and Time* (*Sein und Zeit*) from 1927 is much discussed and has had a broad impact that can hardly be underestimated (cf. Figal 2000. 192–209; Steinmann 2010. 103–110). In the following, we will therefore only interpret Heidegger's concept of *Angst* to the extent that this is necessary to understand the Jonasian approach in its connection to and, at the same time, its demarcation from Heidegger.

Heidegger addresses the phenomenon of *Angst* in *Being and Time* when he asks about the "structural whole of the everydayness of Da-sein in its totality" (Heidegger 1996. 170; Heidegger 1977. 241). In terms of content, Heidegger determines this wholeness of human *Dasein* through the structure of care (*Sorge*). But this must first be shown phenomenologically or made tangible through a phenomenon, "in which Da-sein brings itself before itself", in such a way „that in it Da-sein becomes accessible to itself, so to speak, in a *simplified way*" (Heidegger 1996. 170; Heidegger 1977. 242). According to Heidegger, this phenomenon is *Angst* as a fundamental kind of attunement (*Grundbefindlichkeit*), which always perceives *Dasein* explicitly or implicitly in its finiteness. In order to un-

<sup>2</sup> Jonas studied with Heidegger during his time in Marburg, as evidenced by the Schelling Seminar that Jonas attended in 1927–1928, which also dealt with the concept of "*Angst*" (cf. Heidegger 2010. 291, 311, 314 and 344).

derstand what the feeling of *Angst* is supposed to capture phenomenally here, it is worthwhile first of all to briefly visualize the structure of *Angst*, before we can ask how this structure becomes present in a feeling or attunement.

Heidegger understands care as the being of human *Dasein*. This “lies ‘before’ every factual ‘attitude’ and ‘position’ of Da-sein, that is, it is always already *in* them as an existential *a priori*” (Heidegger 1996. 180; Heidegger 1977. 257). Therefore, care should not be equated with special acts or drives like wanting and desires (*Wünschen*) or urge (*Drang*) and predilection (*Hang*). In these everyday behaviors, the underlying care structure is no longer present in its entirety, or only in a modified way (cf. Höfele 2019. 299–304). In its entirety, the structure of care is characterized by three essential moments that, according to Heidegger, describe the being of *Dasein* ontologically: In recourse to the stoic tradition of *cura sui*, Heidegger defines *Dasein* as a being that is concerned with itself. In this way, however, *Dasein* is always already free and open “for its ownmost potentiality-for-being (*für das eigenste Seinkönnen*)”, which it has to grasp and concretize; as such, a “being-ahead-of-itself (*Sich-vorweg-sein*)” characterizes *Dasein* in general, which already hints at the *future orientation* of *Dasein* that lies in ability (*Können*) (Heidegger 1996. 179; Heidegger 1977. 254 f.). But the “existing is always factual”, as Heidegger adds with regard to the second moment of the care structure (Heidegger 1996. 179; Heidegger 1977. 255). *Dasein* is always already “thrown” (*geworfen*) into a world that is given to it as already *having been* (*gewesen*). Nevertheless, *Dasein* does not find itself here as an isolated subject placed in the world, but the “thrown potentiality-for-being-in-the-world (*geworfenes In-der-Welt-sein-können*) [...] is always already also absorbed in the world taken care of” (Heidegger 1996. 179; Heidegger 1977. 255). It is always already in the *presence* of being encountered, which it deals with every day.

In its tripartite nature, care refers to “*Gewesenheit*” (past), present and future – thus to the three dimensions of temporality (*Zeitlichkeit*), which constitute and guarantee the “*primordial unity*” of care (Heidegger 1996. 301; Heidegger 1977. 433). It is only with the development of the three structural moments of care that Heidegger can understand these moments as a “being-ahead-of-oneself-already-in (the world) as being-together-with (innerworldly beings encountered) (*Sich-vorweg-schon-sein-in-(der-Welt-) als Sein-bei (innerweltlich belegendem Seienden)*”. That is to say, they form a uniform and fundamental structure, which is to be distinguished from all purely ontic phenomena “as worry or carefreeness” (*Besorgnis, bzw. Sorglosigkeit*), insofar as it is the basis for them (Heidegger 1996. 180; Heidegger 1977. 256).

*Angst* has to show this threefold structure of care in a phenomenal way. It has to facilitate the experience that *Dasein* as being-in-the-world (*In-der-Welt-sein*) is always already open for the realization of its own *future* possibilities on the ground of its *factual* existence in a *present*. But since care is not an everyday concrete behavior, but always already implicitly characterizes *Dasein*, Heidegger

feels methodically compelled to assert the care structure and the being-in-the-world of *Dasein ex negativo*. Already in the analysis of the handiness (*Zuhandenheit*) of inner-worldly existing stuff in § 16, Heidegger remarks that its essence can be discovered precisely through its unusability and unavailability. For it is only through this that its otherwise always unthematic handiness comes to light. In its unusability, handiness "does not just disappear, but bids farewell, so to speak, in the conspicuousness of what is unusable. Handiness shows itself once again, and precisely in doing so the worldly character of what is at hand also shows itself, too." (Heidegger 1996. 69; Heidegger 1977. 100.) Analogously, *Angst* is also supposed to illustrate *Dasein* as being-in-the-world, just as being-in-the-world gets lost and becomes uncanny for the *Dasein* in the world, in that it just does not feel at home: "In *Angst* one has an 'uncanny' feeling. [...] But uncanniness (*Unheimlichkeit*) means at the same time not-being-at-home" (Heidegger 1996. 176; Heidegger 1977. 250) But this feeling of not being at home not only leads, according to Heidegger, to an isolation of *Dasein*. At the same time, it reveals to *Dasein* its authentic being, which it has implicitly always been in being-in-the-world: namely "*being toward* its ownmost potentiality of being, that is, *being free for* the freedom of choosing and grasping itself (*Freisein für die Freiheit des Sich-selbst-wählens und -ergreifens*)"; *Angst* shows *Dasein* „the authenticity of its being as possibility which it always already is" (Heidegger 1996. 176; Heidegger 1977. 249 f.).

According to Heidegger, *Angst* makes it obvious that the open character and freedom of *Dasein* is something that must be wrested from the openness of the future of *Dasein*. Only when this openness is ensured, can freedom happen to *Dasein*, thus enabling *Dasein* to essentially comprehend itself in its open character (cf. Steinmann 2010. 103–110). This is precisely the point that Jonas makes when he sees modern technology and its dangers as calling into question not only the "human fate" (*Menschenlos*) but also the "human image" (*Menschenbild*). Jonas' concern to preserve that which man is refers precisely to this horizon of possibility of man, which is stretched out in the temporality of *Dasein*, as the revelation of Jonas' connection to Heidegger makes even clearer.<sup>3</sup> In Jonas' eyes, the danger of man's technical actions in the present consists precisely in robbing future generations of this horizon of possibility, or at least restricting it, and thus substantially curtailing their humanity.

<sup>3</sup> As Heidegger's later philosophy of technology shows, Heidegger also sees in technology a danger for the human image, namely that it is "only taken for continuance itself (*selber nur noch als Bestand genommen*)" (Heidegger 2000. 28). In *Being and Time*, Heidegger has already indicated that the structure of care can be related to the technical action of man, at least indirectly, by referring to Goethe's *Faust II* (Heidegger 1996. 405 n. 5 / Heidegger 1977. 262 n. 1), where in the fifth act the "technician" Faust is haunted by the personified Care (*Sorge*) against the background of his ruthless land reclamation project (Goethe, *Faust II*, vv. 11382 ff.).

That Jonas, however, does not adopt Heidegger's *Angst* analysis unchanged, is already shown by the fact that instead of *Angst* he speaks almost exclusively of fear (*Furcht*), which according to Heidegger is ontologically subordinate to the former. Despite his commitment to Heidegger's approach, Jonas has two points of criticism with regard to Heidegger.

(1) Particularly with regard to the function of his "heuristics of fear", namely to act as a mediating instrument between ethical knowledge of principles and their political application, Jonas cannot go along with the Heideggerian determination of the object of *Angst*. According to Heidegger, the fundamental attunement of *Angst* is characterized by the fact that it cannot precisely name what *Angst* is about: "The fact that what is threatening is *nowhere* characterizes what *Angst* is about" (Heidegger 1996. 174; Heidegger 1977. 248). This does not mean, however, that *Angst* is afraid of a mere chimera and is therefore unfounded. "But 'nowhere' does not mean nothing; rather, region in general lies therein, and disclosedness (*Erschlossenheit*) of the world in general for essentially spatial being-in (*In-Sein*)" (Heidegger 1996. 174; Heidegger 1977. 248). According to Heidegger, the indeterminacy of *Angst* makes it possible to bring the whole of the world into view:

The utter insignificance which makes itself known in the nothing and nowhere does not signify the absence of world, but means that innerworldly beings in themselves are so completely unimportant that, on the basis of this *insignificance* of what is innerworldly, the world is all that obtrudes itself in its worldliness (Heidegger 1996. 175; Heidegger 1977. 248).

Since no specific inner-worldly being is the focus of *Angst*, only the whole, namely the worldliness of the world, can come into view. This is precisely the difference to fear (*Furcht*), which Heidegger therefore subordinates to *Angst* ontologically. Fear always has a concrete what-about (*Wovor*), as Heidegger previously explained in § 30: "That before *which* we are afraid (*Wovor der Furcht*), the 'fearsome,' is always something encountered within the world, either with the kind of being of something at hand (*Zuhandenen*) or something objectively present (*Vorhandenen*) or *Mitda-sein*" (Heidegger 1996. 131 f.; Heidegger 1977. 186; cf. also Figal 2000. 195; Steinmann 2010. 107). Since technical means are also to be subsumed under what is at hand (*Zuhandenes*), this moment of fear is understandably more interesting for Jonas's technical-ethical approach than the indeterminacy of the object of *Angst*. In its application-relatedness, Jonasian ethics can connect to this moment of fear, especially since Heidegger refers to the rational, clarifying moment in which fear exists: "And then fear, in being afraid, can 'clarify' what is fearsome by explicitly looking at it" (Heidegger 1996. 132; Heidegger 1977. 187). At the same time, however, Jonas does not want to renounce the principle-theoretical significance of *Angst* and its openness to the

possibility character of *Dasein*, which is why he does not exclusively insist on the notion of fear (*Furcht*) in *The Imperative of Responsibility* (cf. Jonas 1979. 392).

(2) Nevertheless, Jonas insists that he means "in no way fear or anxiety for oneself (*Furcht oder Angst um sich selbst*)" (Jonas 1979. 392), which in a second point distinguishes him from Heidegger. For, according to Heidegger, both that for which (*worum*) fear is afraid and that for which *Angst* is afraid are *Dasein* itself as being-in-the-world (cf. Heidegger 1996. 132, 175f.; Heidegger 1977. 188, 249): "*Angst* individuates Da-sein to its ownmost being-in-the-world which, as understanding, projects itself essentially upon possibilities" (cf. Heidegger 1996. 176; Heidegger 1977. 188, 249). In this way Heidegger seeks to secure the very authenticity (*Eigentlichkeit*) of *Dasein*, which he endeavors to delimit from the falling prey (*Verfallenheit*) to the collective, individuality-free They (*Man*).

Jonas, however, is concerned precisely with delimiting "the self-committing freedom of the self", and in this he sees himself precisely in contrast to "Heidegger's 'resoluteness,' [...] where the worldly issue is not by itself endowed with a claim on us but receives its significance from the choice of our passionate concern." (Jonas 1984. 88; Jonas 1979. 167.) In this respect, he is also aiming precisely at the feeling of a "selfless fear" (Jonas 1979. 392; cf. Jonas 1984. 162), which is unthinkable for Heidegger. But Jonas wants to ensure with it the inclusion of future generations of mankind in that for which fear is afraid. True humanity or the ethically responsible conception of man is not guaranteed for him if man as an individual reflects on himself, but only if he includes the future other in this self-reflection.

With this motive, however, that true humanity or self-consciousness can only be achieved through the mediation of another, Jonas approaches another philosopher, namely G. W. F. Hegel, to the same extent that he distances himself from Heidegger. This is all the more astonishing because in *The Imperative of Responsibility* Jonas seeks to distance himself from Hegel's eschatological philosophy of history, which was continued by Marx and the Marxism that Jonas criticizes in the person of Ernst Bloch (cf. Jonas 1984. 127; Jonas 1979. 228). Without Hegel being named in this respect, Jonas seems to go back to one of the central roots of Marxist theory (cf. e.g. Kojève 1947; Althusser 1976) in order to make it fruitful for his ethical approach in an implicit follow-up, namely to the chapter "Independence and dependence of self-consciousness: lordship and bondage" from the Hegelian *Phenomenology of Spirit* of 1807 (Hegel 2018. 76–82; Hegel 1980. 109–116). For Hegel, too, the feeling of fear is central in the struggle between lord and bondsman, which he also links to the motif of selflessness in a similar way to Jonas.



#### IV. THE FEAR OF THE ENSLAVED MANUFACTURER AS A CONDITION FOR A NEW LEVEL OF CONSCIOUSNESS

The chapter “Lordship and Bondage” leads consciousness towards self-consciousness. But the level of self-consciousness is only reached through the mutual recognition of two self-consciousnesses, as Hegel states in the tradition of the Fichtean theory of recognition (cf. Honneth 1992. 11–106): “There is a *self-consciousness for a self-consciousness*. Only so is it in fact self-consciousness; for only in this way does the unity of itself in its otherness come to be for it” (Hegel 2018. 76; Hegel 1980. 108). Alexandre Kojève saw the decisive difference between Hegel’s and Heidegger’s conception of true self-consciousness or authentic *Dasein* in the fact that, according to Heidegger, this only occurs in isolation, whereas according to Hegel it is only possible intersubjectively. In particular, this difference between the two thinkers becomes apparent in the role of the feeling of fear in the struggle for recognition in Hegel’s work, which in the case of Hegel does not spring from a passive contemplation of being-to-death as in Heidegger’s work, but from the active negation of the other.<sup>4</sup>

In order to understand the role of this feeling of fear in the Hegelian conception of the struggle for recognition and the selflessness it involves, it is necessary to trace at least the decisive steps in this much-interpreted movement for recognition (cf. e.g. Gadamer 1987; Siep 2014. 90–95; Stekeler 2014. 663–719). In contrast to the preceding concepts of recognition, especially those of Fichte, but also Schelling (cf. Höfele 2019. 22 f., 37–44, 71), Hegel describes an asymmetry between the two self-consciousnesses (as lord and bondsman) in the movement of recognition. On the other hand, he gives it a historical-existential meaning by speaking of a struggle that could then be reinterpreted as the history of class struggle, especially in the later Marxism of the 20th century.

Hegel begins the chapter still completely in the sense of the movement of recognition established by Fichte: “Self-consciousness is *in* and *for itself*, when, and by the fact that, it is in and for itself for another self-consciousness; that is, it is only as something recognized” (Hegel 2018. 76; Hegel 1980. 109). But, as for self-consciousness there is another self-consciousness through which it looks at itself, it has on the one hand lost itself in this other, and at the same time has suspended the other, in so far as it regards this other only as a mirror of itself. Thus this double suspension of the two self-consciousnesses must be followed by “an

<sup>4</sup> Cf. Kojève 1993. 39: “Seulement, à l’encontre de Heidegger, Hegel affirme que ce n’est pas l’angoisse de la contemplation passive de l’approche de sa fin biologique, mais uniquement l’angoisse dans et par la lutte pour la mort, c’est-à-dire dans et par la *négation*-active de l’être donné comme un Ce-qui-est-comme-lui-sans-être-lui (bref : d’un autre homme), d’un être qui peut ainsi le nier activement lui-même, que c’est seulement la mort révélée dans et par cette lutte négatrice qui a la valeur humaine ou – plus exactement – humanisante que lui attribue Heidegger.” Cf. to the fear in Hegel also Gretic 2002.



ambiguous return into *itself*", whereby both self-consciousnesses are given back to themselves (Hegel 2018. 77; Hegel 1980. 109).

But this ideal or pure movement of recognition, which is to be carried out by both self-consciousnesses in the indicated way, cannot take place in perfect parallelism according to Hegel, and thus "the process will present the side of the *inequality* of the two, or the bifurcation of the middle term into the extremes which, as extremes, are opposed to one another, one being only recognized, the other only recognizing" (Hegel 2018. 77; Hegel 1980. 110). The intention of suspending the respective other's self-consciousness is existentially understood as "a life-and-death combat" (Hegel 2018. 78; Hegel 1980. 111). The goal of the fight must not be death as a complete abstraction of one of the two self-consciousnesses, which would make self-recognition in the other impossible, as in the case of desire consuming the object. The struggle must therefore lead to an "abstract negation, not the negation of consciousness, which *sublates* in such a way as to *preserve* and *maintain* what is sublated, and thereby survives its being sublated" (Hegel 2018. 79; Hegel 1980. 112).

The result of that struggle must therefore lie in the establishment of two gradually different self-consciousnesses. The one self-consciousness must be able to be completely for itself as master and winner of the struggle, while the inferior self-consciousness as bondsman is committed to one being-for-another, namely for the Lord: "one is the independent consciousness for which the essence is Being-for-itself, the other is the dependent consciousness for which the essence is life or Being for another; the former is the *lord*, the latter is the *bondsman*" (Hegel 2018. 79; Hegel 1980. 112).

The Lord can thus enjoy the object of his desire mediated by the bondsman who works it. The bondsman, on the other hand, must deal with the indissoluble materiality of the object, which his master may therefore perceive and consume as a pure object of pleasure. The bondsman, however, "cannot through his negating have done with it to the point of annihilation, or he only *works on* it" (Hegel 2018. 79; Hegel 1980. 113). But here, a dialectical movement asserts itself simultaneously. For the Lord looks upon his own otherness in the servant as the unessential consciousness: „The *truth* of the independent consciousness is accordingly the *servile consciousness*" (Hegel 2018. 80; Hegel 1980. 114). The bondsman, by contrast, finds in the Lord his object as his own other, while at the same time, he possesses an independent being that goes back to himself in the indissoluble materiality on which he imprints his own form: "the working consciousness arrives at the intuition of independent Being *as of its own self*" (Hegel 2018. 81; Hegel 1980. 115). In this respect, it is precisely the servile self-consciousness that, according to Hegel, achieves a "new shape of self-consciousness" (Hegel 2018. 82; Hegel 1980. 116).

But the bondsman achieves this new form of self-consciousness through three moments, of which one essential moment is fear:

[1] In the lord, the Being-for-itself is *an other* to it, or only *for it*; [2] in fear (*Furcht*), the Being-for-itself is *within itself*; [3] in its cultivating, the Being-for-itself becomes for it as *its own* Being-for-itself, and it arrives at the consciousness that it itself is in and for itself” (Hegel 2018. 81; Hegel 1980. 115; trans modified).

(1) In the Lord, who in the struggle has risen to independent self-consciousness, the bondsman is confronted with his being-for-himself in the form of another. Paradoxically, this independence could only be revealed to him in that he humiliated himself, placing himself in the position of a dependent self-consciousness and acting selflessly. But this moment, as such, would only result in the self-loss of servile self-consciousness mentioned at the beginning. (2) Therefore, the fear of the second moment is essential, as it makes the bondsman’s being-for-himself experienceable in himself. As Hegel remarks emphatically before, this is not a momentary fear, but a fear that seizes the whole essence:<sup>5</sup>

In this [fear of death] it has been internally dissolved, has trembled through and through within itself, and everything fixed has quaked in it. But this pure universal movement, the absolute liquidization of all subsistence, is the simple essence of self-consciousness, absolute negativity, *pure Being-for-itself*, which is thus *within* this consciousness. (Hegel 2018. 80; Hegel 1980. 114.)

Only here with fear is the central determination of self-consciousness achieved, namely, that it is absolute negativity, i.e. a structure, which, like Heidegger’s open character of *Dasein*, has nothing static about it, but is pure motion. Only in this liquefaction of consciousness, which no longer knows any substantial peculiarity, is it possible for it to take on a higher level or a new form. In this respect, one can speak here with Jonas of a form of “selfless fear”, since this fear lets the servile self-consciousness become selfless in relation to the Lord on the one hand, and, on the other hand, it opens the bondsman to future forms of consciousness that are still to be attained. Even if it is unclear whether Jonas had this chapter of Hegelian phenomenology in mind, this motif shows a striking parallel to Jonas’ “heuristics of fear”. Although Hegel does not explicitly have future generations in mind, Hegel’s concept of self-consciousness encompasses a plurality of subjects, namely “*I* that is *We*, and *We* that is *I*” (Hegel 2018. 76; Hegel 1980. 108), and not just a particularity as in the case of Heidegger’s *Dasein*.

(3) This parallel to Jonas becomes all the more apparent when one considers the third moment of servant self-consciousness, according to which it is precisely in forming and cultivating (technical-craft) that the bondsman becomes aware

<sup>5</sup> Hegel uses the terms “fear” and “anxiety” in exactly the opposite sense, as later Heidegger did: fear applies to the whole being, while anxiety concerns only the individual and is related to this or that moment (cf. Hegel 2018. 80 f.; Hegel 1980. 114 f.).

of his own being-for-himself. Artificial or technical objects are in this respect already classified by Hegel as something that shapes one's own self-image and thus contributes decisively to the "human image" (*Menschenbild*), to quote Jonas. Admittedly, Jonas is by no means striving for a developmental sequence at levels of consciousness like Hegel, and he is more concerned with preserving the essential moments of the "human image" for future generations as well. However, he shares with Hegel the assumption that fear makes a decisive contribution to the self-assurance of human consciousness and that it cannot make this contribution in isolation, but only in the inclusion of a We.

#### V. CONCLUDING CONSIDERATIONS ON THE HOLISTIC CHARACTER OF JONAS' PRECAUTIONARY PRINCIPLE

The dangers that can emanate from new technologies not only affect life and the integrity of humans and the environment. As early as 1979, Jonas saw that technology ethics and technology assessment must take into account not only "human fate" (*Menschenlos*) and its "physical survival" but also the question of whether the "human image" (*Menschenbild*) and the "integrity of [his] essence" are affected by new technologies (Jonas 1979, 8). In a way that already seems to anticipate the discussion points of the Anthropocene debate, Hannah Arendt, another student of Heidegger, also made similar observation. In her book *The Human Condition*, published in 1958, she remarks with regard to modern technology: "The natural processes on which it [the world of machines] feeds increasingly relate it to the biological process itself, so that the apparatuses we once handled freely begin to look as though they were »shells belonging to the human body as the shell belongs to the body of a turtle«" (Arendt 1998, 153; Arendt 1981, 139).<sup>6</sup> Humans seem thereby, as she supplements in the German edition published two years later in 1960, "no longer to belong to the genus of mammals, but begin to transform themselves into a kind of shellfish" (*beginne sich in eine Art Schaltier zu verwandeln*; Arendt 1981, 139).

Against the background of this insight, Jonas, by means of his "heuristics of fear", thus sought to achieve more than is envisaged by the precautionary principle formulated in § 15 of the *Report of the United Nations Conference on Environment and Development*. According to the report, the "lack of full scientific certainty shall not be used as a reason for postponing cost-effective measures to prevent environmental degradation" (United Nations, 1993, p. 3). It is true that the "heuristics of fear" also has the task of assessing the risk of technical developments and, based on this, of focusing on application-oriented measures. This is probably one of the main reasons why Jonas, despite his obvious connection to

<sup>6</sup> Here Arendt quotes Werner Heisenberg's *Das Naturbild der heutigen Physik* from 1955.

Heidegger's analysis of anxiety, speaks of a "heuristics of *fear (Furcht)*" and not of "*Angst*". For fear always goes back to a concrete cause or a certain object, to which it is necessary to react with concrete measures. But with Heidegger, Jonas also has in mind the broader horizon of the orientation of that negative feeling. Fear is intended to bring into view *ex negativo* precisely that which constitutes being human. For technology always has the tendency to inscribe itself like a "shell" into the human being and his life-world and thus to have repercussions on the human being (cf. Ihde 1979; Verbeek 2005). The appropriateness of this tendency must therefore be decided even in view of the lack of knowledge about concrete technical developments.

As could be shown in referring to the concepts serving as precursor to the Jonasian "heuristics of fear", it is by no means a matter of conservatively preserving a static image of humankind. Rather, it is about the preservation of a possibility horizon of *Dasein*. However, Jonas does not only mean the open character of one's own *Dasein* in the sense of Heidegger. Similar to Hegel's concept of fear in "Lordship and Bondage", Jonas is concerned with a form of "selfless fear" that involves a collective We, and that is also worried about future generations and their possibilities. Especially against the background that this idea has not yet been sufficiently considered in ethical technology assessment, Jonas' approach can still be regarded as up-to-date and relevant in this respect.

In doing so, we could even go beyond Jonas to ask to what extent such a "heuristics of fear" can help to consider the impairment of the essence of nature and the environment beyond the reflection on the endangerment of the human being.<sup>7</sup> In view of the interrelationship between humankind, technology and nature, thinking in the Anthropocene is, after all, required to go beyond an anthropocentric point of view, insofar as the latter alone takes into account future generations and the human environment (Höfele 2020). A "heuristics of fear" or "*Angst*" could also be used for a biocentric expansion and could function as a "heuristics for the Anthropocene", insofar as fear is an emotion that is not only inherent in humans but also in animals (Soentgen 2018) and thus also allows for reflection on their essential endangerment in the Anthropocene.

<sup>7</sup> Although Hans Jonas dealt extensively with the philosophy of nature in *The Phenomenon of Life: Toward a Philosophical Biology*, nature is only insufficiently considered from an anthropocentric perspective in *The Imperative of Responsibility* (cf. Jonas 1984. 186 ff.; Jonas 1979. 327 ff.).

## REFERENCES

- Althusser, Louis 1976. *Idéologie et appareils idéologiques d'État*. (Notes pour une recherche). In id. *Positions (1964–1975)*. Paris, Les Éditions sociales [1st ed. 1970]. 67–125.
- Arendt, Hannah 1981. *Vita activa oder Vom tätigen Leben*. München–Zürich, Piper [1st ed. 1960].
- Arendt, Hannah 1998. *The Human Condition*. 2nd ed. Chicago, The University of Chicago Press [1st ed. 1958].
- Böhler, Dietrich 2008. Mitverantwortung für die Menschheitszukunft: Die Aktualität von Hans Jonas. *Osteuropa*. 58/4–5. 20–35.
- Crutzen, Paul J. 2002. Geology of Mankind. *Nature*. 415. 23. <https://doi.org/10.1038/415023a>
- Figal, Günter 2000. *Martin Heidegger. Phänomenologie der Freiheit*. 3rd ed. Weinheim [1st ed. 1988].
- Gadamer, Hans-Georg 1987. Die Dialektik des Selbstbewußtseins. In id. *Gesammelte Werke*. Vol. 3: *Neuere Philosophie I. Hegel, Husserl, Heidegger*. Tübingen, J.C.B. Mohr [1st ed. 1973]. 47–64.
- Gretic, Goran 2002. Freiheit, Schrecken und Angst. *Hegel-Jahrbuch 2002*. 143–146.
- Grunwald, Armin 2002. *Technikfolgenabschätzung – eine Einführung*. Berlin, Edition Sigma.
- Hegel, Georg Wilhelm Friedrich 1800. *Gesammelte Werke*. Ed. by the Nordrhein-Westfälische Akademie der Wissenschaften und der Künste. Vol. 9: *Phänomenologie des Geistes*. Ed. by Wolfgang Bonsiepen and Reinhard Heede. Hamburg, Felix Meiner Verlag [1st ed. 1807].
- Hegel, Georg Wilhelm Friedrich 2018. *The Phenomenology of Spirit*. Translated with introduction and commentary by Michael Inwood. Oxford, Oxford University Press.
- Heidegger, Martin 1977. *Gesamtausgabe*. Vol. 2: *Sein und Zeit*. Ed. by Friedrich-Wilhelm von Herrmann. Frankfurt/M, Vittorio Klostermann [1st ed. 1927].
- Heidegger, Martin 1996. *Being and Time. A Translation of Sein und Zeit*. Trans. by Joan Stambaugh. Albany/NY, State University of New York Press.
- Heidegger, Martin 2000. Die Frage nach der Technik. In id. *Gesamtausgabe*. Vol. 7. Ed. by Friedrich-Wilhelm von Herrmann. Frankfurt am Main, Vittorio Klostermann [1st ed. 1953]. 5–36.
- Heidegger, Martin 2010. Protokolle einer Übung von Martin Heidegger zu 'Schellings Abhandlung über das Wesen der menschlichen Freiheit' aus dem Wintersemester 1927/28 in Marburg. Ed. by Philipp Schwab – Sebastian Schwenzfeuer. In Lore Hühn – Jörg Jantzen (eds.) *Heideggers Schelling-Seminar (1927/28). Die Protokolle von Martin Heideggers Seminar zu Schellings „Freiheitschrift“ (1927/28) und die Akten des Internationalen Schelling-Tags 2006*. Stuttgart – Bad Cannstatt, Frommann-Holzboog. 263–463.
- Höfele, Philipp 2019. *Wollen und Lassen. Zur Ausdifferenzierung, Kritik und Rezeption des Willensparadigmas in der Philosophie Schellings*. Freiburg – München, Verlag Karl Alber.
- Höfele, Philipp 2020. The Changed Role of Anthropology in the Anthropocene. In Silvia Salardi – Michele Saporiti (eds.) *Le tecnologie 'moralì' emergenti e le sfide etico-giuridiche delle nuove soggettività. Emerging 'Moral' Technologies and the Ethical-Legal Challenges of New Subjectivities*. Turin, Giappichelli. 125–144.
- Honneth, Axel 1992. *Kampf um Anerkennung. Zur moralischen Grammatik sozialer Konflikte*. Frankfurt/M, Suhrkamp.
- Huber, Wolfgang 2018: Ehrfurcht vor dem Heiligen. Zur Aktualität des *Prinzip Verantwortung* von Hans Jonas. *Pastoraltheologie*. 107. 411–426.
- Ihde, Don 1979: *Technics and Praxis*. Dordrecht–Boston–London, D. Reidel Publishing Company.
- Jonas, Hans 1979. *Das Prinzip Verantwortung. Versuch einer Ethik für die technologische Zivilisation*. Frankfurt/M, Suhrkamp.

- Jonas, Hans 1984. *The Imperative of Responsibility: In Search of Ethics for the Technological Age*. Translated by Hans Jonas with the Collaboration of David Herr. Chicago, University of Chicago Press.
- Jonas, Hans 2001. *The Phenomenon of Life: Toward a Philosophical Biology*. With a foreword by Lawrence Vogel. Evanston, Northwestern University Press [1st ed. 1966].
- Kant, Immanuel 1900ff. *Gesammelte Schriften*. Ed. by the Berlin-Brandenburgische Akademie der Wissenschaften. Berlin, De Gruyter [= AA].
- Kojève, Alexandre 1947. *Introduction à la lecture de Hegel. Leçons sur la Phénoménologie de l'esprit professées de 1933 à 1939 à l'École des Hautes Études*. Ed. by Raymond Queneau. Paris, Gallimard.
- Kojève, Alexandre 1993. Note sur Hegel et Heidegger. *Logiques de l'éthique*. 7. 35–46.
- Lenton, Timothy M. et al. 2019. Climate Tipping Points – Too Risky To Bet Against. *Nature*. 575. 592–595 <https://doi.org/10.1038/d41586-019-03595-0>
- von Sass, Hartmut 2016. Heuristik der Furcht. Endlosigkeit als kontrafaktische Unterstellung – oder: selbst Hans Jonas hofft, aus Prinzip. *Hermeneutische Blätter*. 1. 161–176.
- Schmidt, Jan C. 2013. Prinzip Verantwortung. In Armin Grunwald (ed.) *Handbuch Technikethik*. Stuttgart–Weimar, J.B. Metzler. 143–148.
- Siep, Ludwig 2014. *Hegel's Phenomenology of Spirit*. Transl. by Daniel Smyth. Cambridge, Cambridge University Press [1st ed. 2000].
- Soentgen, Jens 2018. *Ökologie der Angst*. Berlin, Matthes & Seitz.
- Steffen, Will [et al.] 2015. The Trajectory of the Anthropocene: The Great Acceleration. *The Anthropocene Review*. 2/1. 81–98. <https://doi.org/10.1177/2053019614564785>
- Steinmann, Michael 2010. *Martin Heideggers 'Sein und Zeit'*. Darmstadt, WBG.
- Stekeler, Pirmin 2014. *Hegels Phänomenologie des Geistes. Ein dialogischer Kommentar*. Vol. 1. *Gewissheit und Vernunft*. Hamburg, Felix Meiner Verlag.
- United Nations 1993. *Report of the United Nations Conference on Environment and Development*. Vol. 1: *Resolutions adopted by the conference*. New York, United Nations.
- Verbeek, Peter-Paul 2005. *What Things Do: Philosophical Reflections on Technology, Agency, and Design*. University Park, The Pennsylvania State University Press.
- Zalasiewicz, Jan et al. 2017. Scale and Diversity of the Physical Technosphere: A Geological Perspective. *The Anthropocene Review*. 4/1. 9–22. <https://doi.org/10.1177/2053019616677743>

## Contributors

GERGELY AMBRUS is Associate Professor of Philosophy at Eötvös Loránd University, Budapest. His main areas of specialisation are: 20<sup>th</sup> century philosophy of mind and the history of analytic philosophy with and emphasis on their intersection.

E-mail: [ambrus.gergely@btk.elte.hu](mailto:ambrus.gergely@btk.elte.hu)

GÁBOR BOROS is Professor of Philosophy at Károli Gáspár University of the Reformed Church in Hungary. His main areas of specialisations are early modern philosophy, philosophy of emotions – especially love –, and practical philosophy in general.

E-mail: [h13556bor@mail.iif.hu](mailto:h13556bor@mail.iif.hu)

PHILIPPE CABESTAN (PhD) teaches philosophy at Lycée Janson de Sailly Paris and is President of the French Society of Daseinsanalyse. His research areas include phenomenology, Sartre, anthropology and psychoanalysis.

E-mail: [philippecabestan@orange.fr](mailto:philippecabestan@orange.fr)

JAMES CARTLIDGE is a PhD Student at Central European University where he is working on his thesis about Heidegger's philosophical anthropology.

E-mail: [jkcartlidge92@gmail.com](mailto:jkcartlidge92@gmail.com)

PHILIPP HÖFELE (PhD) is Postdoc Researcher at University of Freiburg. He was educated at University of Freiburg, Paris-Sorbonne University and at University of Strasbourg. His fields of interest include ethics, philosophy of nature and technology, aesthetics, philosophy of religion and German idealism.

E-mail: [philipp.hoefele@philosophie.uni-freiburg.de](mailto:philipp.hoefele@philosophie.uni-freiburg.de)

LORE HÜHN (PhD) is Professor of Philosophy specializing in Ethics and German Idealism at the University of Freiburg. She was educated at Marburg University and at Freie Universität Berlin. Her research interests include ethics, aesthetics, philosophy of nature, anthropology, German idealism and philosophy of the 19<sup>th</sup> and 20<sup>th</sup> century.

E-mail: [lore.huehn@philosophie.uni-freiburg.de](mailto:lore.huehn@philosophie.uni-freiburg.de)

**HYE YOUNG KIM (PhD)** is Associate Researcher at the Husserl Archive at l'Ecole Normale Supérieure in Paris. She was educated Korea University (South Korea) and Freie Universität Berlin. Her fields of interest are 19th and 20th century European philosophy, metaphysics, epistemology, foundational mathematics.

E-mail: [hye.young.kim@ens.fr](mailto:hye.young.kim@ens.fr)

**CSABA OLAY (PhD)** is Professor of Philosophy at Eötvös University Budapest. He was educated at Eötvös University and at University of Freiburg. His fields of interest include 19th and 20th century continental philosophy, hermeneutics, existentialism, aesthetics.

E-mail: [olay.csaba@btk.elte.hu](mailto:olay.csaba@btk.elte.hu)

**CSABA PLÉH** is distinguished visiting professor of the Central European University, Department of Cognitive Science, Fellow of the Collegium de Lyon, Université de Lyon, full member of the Hungarian Academy of Sciences. His interests: cognitive psychology, history of cognition, theories of psychology.

E-mail: [vispleh@ceu.edu](mailto:vispleh@ceu.edu)

**TIM THORNTON (PhD, DLitt)** is Professor of Philosophy and Mental Health at the University of Central Lancashire. He was educated at Cambridge University. His research interests are philosophy of psychiatry, judgement, tacit knowledge and the philosophy of Wittgenstein and of John McDowell.

E-mail: [Tthornton1@uclan.ac.uk](mailto:Tthornton1@uclan.ac.uk)

**JUDIT SZALAI** is Associate Professor of Philosophy at Eötvös University, Budapest. Her main interests are philosophical psychology and psychopathology, the philosophy of mind, the philosophy of technology, and the history of seventeenth-century thought.

E-mail: [szalai@caesar.elte.hu](mailto:szalai@caesar.elte.hu)

**DAVID WEBERMAN (PhD)** is Associate Professor of Philosophy at the Central European University. He was educated at Ludwig-Maximilians-Universität Munich and Columbia University. His research interests include 19th and 20th century European philosophy; philosophy of history; hermeneutics; aesthetics.

E-mail: [webermand@ceu.edu](mailto:webermand@ceu.edu)



# Summaries

GERGELY AMBRUS

## **Personal Identity: Narrative Constitution vs Psychological Continuity**

The paper investigates Marya Schechtman's narrativist account of self and personal identity, which she dubbed the "Narrative Self-Constitution View". I lay out the main features of this conception by contrasting it with the views of Derek Parfit, a major contemporary representative of the psychological relationalist tradition originating from Locke and Hume, to which Schechtman's theory, and narrativism in general, may be seen as a major alternative. After presenting the main features of these two accounts, I set out to defend Parfit against an important criticism of Schechtman which seeks to discredit Parfit's notion of quasi-memory (and quasi-belief, quasi-desire etc. as well). Parfit's psychological continuity view essentially depends on these notions, hence undermining them provides a ground for accepting narrativism. However, I also argue that the psychological continuity view fails seriously as well, as it does not account for identification I take to be a necessary condition of being the same person. Lastly, I discuss certain possible explanations of identification, and address the question whether these support the narrativist or the psychological continuity view.

*Keywords:* narrative constitution, psychological continuity, Schechtman, Parfit, q-memory.

GÁBOR BOROS

## **First Phase of the Narrative Theory of Identity and Emotions: Dilthey, Misch**

In the last decades, a number of philosophical and psychological theories have made serious attempts to discover and make use of various aspects of different types of narratives from the "novels of time" through biographies and autobiographies to interviews with members of contemporary groups or individuals in therapeutic analysis. Their aims were to understand identity in a broad sense, personal, group-, national, emotional, and other types of it. The initiators and proponents of these theories rarely referred to the two German philosophers, active between the mid-19<sup>th</sup> and the mid-20<sup>th</sup> century who attributed a fundamental role to autobiography as a particular kind of narrative both in history and in philosophy. In spite of their being neglected, Wilhelm Dilthey and Georg Misch merit to be integrated in the history of the narrative identity-movement. They emphasised the fundamental importance of establishing narrative connections between the seemingly isolated events in life histories as the most effective instrument to establish meaningful

and coherent life-units. This paper will regard Dilthey and Misch as our contemporaries in order to weigh up their contributions to a renewal and enrichment of the theory of narrative identity and the narrative theories of emotions. Our other aim is to persuade the reader that from this perspective, it is worth involving them in the general discourse on identity and narrativity.

*Keywords:* autobiography, narrative identity, Georg Misch, Wilhelm Dilthey

PHILIPPE CABESTAN

### **Bad Faith *versus* Unconscious: a Credible Alternative?**

It must be admitted that Freud's way of thinking has already lost a part of its credit. However, it is certainly not a reason to claim that the hypothesis of the unconscious is null and void. Therefore, in the paper I will examine whether and how it's possible to keep Freud's hypothesis. First, I would like to draw attention to Heidegger's criticism, which is based on an ontological argument, according to which natural and human phenomena belong to different forms of being. Consequently, one cannot speak of human behaviour as if it were a physical or chemical process. Then, I shall discuss the relation of Sartre's concept of bad faith to the notion of unconscious behaviour. Sartre, as illustrating his concept, proposes that a hysteric is aware of what he doesn't want to be aware of; and, as long as he tries to escape from it, he is necessarily aware of what he tries to escape from. But, of course, the explanations of unconscious behaviour which rely on the concept of bad faith cannot be exhaustive; furthermore, we also have to make room for a notion of unconscious which does not imply that we consider the unconscious as a noun (MacIntyre) or as a thing in-itself (Sartre).

*Keywords:* unconscious, bad faith, Sartre, Freud.

JAMES CARTLIDGE

### **Heidegger's Philosophical Anthropology of Moods**

Martin Heidegger often and emphatically claimed that his work, especially in his masterpiece *Being and Time*, was not philosophical anthropology. He conceived of his project as 'fundamental ontology', and argued that because it is singularly concerned with the question of the meaning of Being in general (and not 'human being'), this precluded him from being engaged in philosophical anthropology. This is a claim we should find puzzling because at the very heart of Heidegger's project is an analysis of the structures of the existence of 'Dasein', an entity that human beings are an instantiation of, the entity that has a relationship of concern towards its existence and which is capable of raising the question of the meaning of Being. Heidegger argues that because he only analyses human beings insofar as they are Daseins, he cannot be doing philosophical anthropology, but only fundamental ontology. In this paper, I refute this claim. I provide a sketch of philosophical anthropology which conceives it as the other side of anthropology's coin. Where anthropology is concerned with understanding human difference, philosophical anthropology attempts to understand what is common to all instances of human existence and elucidate its significant features and structures. Whether he likes it or not,

Heidegger is engaged in exactly this kind of project because *anything that applies truly to Dasein applies truly to human beings*. With this in mind, I examine Heidegger's analysis of moods to demonstrate that Heidegger's work is best understood as involving a kind of philosophical anthropology.

*Keywords:* Heidegger, Continental Philosophy, Metaphysics, Ontology, Anthropology, Phenomenology, Existentialism, Moods, Emotions.

PHILIPP HÖFELE

### **New Technologies and the 'Heuristics of Fear': The Meaning and Prehistory of an Emotion in Jonas, Heidegger and Hegel**

In view of the ever-increasing dominance of technology in modernity, which goes hand in hand with an ever-faster development of new technologies as well as an ever-increasing blurring of the boundaries between the natural and the artificial, an evaluation of emerging technologies cannot be put off. As early as 1979, in his work *The Imperative of Responsibility*, Hans Jonas developed a 'heuristics of fear' as a principle and method for assessing the dangers of new technologies in the face of ignorance of their future consequences. As the article attempts to show, the task of this 'heuristics' amounts to more than assessing the risk of technical developments and presenting application-oriented measures based on this assessment. As a feeling, the fear should also illustrate *ex negativo* that which constitutes human existence as such, insofar as technology always inscribes itself in the self-image of humankind. This becomes clear when one goes back to the historical reference point of Jonas' conception of this feeling in Heidegger's analysis of *Angst*, from which it becomes clear that it is a matter of preserving the horizon of possibilities for human *Dasein*. However, Jonas does not only mean the open character of one's own existence in the sense of Heidegger. Similar to Hegel's concept of fear in "Lordship and Bondage", Jonas is concerned with a form of "selfless fear" that involves a collective We and is also concerned about future generations and their possibilities.

*Keywords:* Jonas, Heidegger, Hegel, technology, Angst, heuristics of fear

LORE HÜHN

### **Com-passion: On the Foundations of Moral Philosophy for J. J. Rousseau and Arthur Schopenhauer**

In a first step, the paper will show how Schopenhauer uses his theory of the essential identity of all living beings, which is founded on his voluntarist metaphysics and orients itself on the criterium of the capacity to suffer, in order to problematize the anthropocentric primacy of reason for the grounding of morals. In a second step, the paper attempts to provide evidence that, despite accepting the negativity of suffering without illusions, it is probable that Schopenhauer's ethics of compassion extends to a horizon of possibility articulated in the medium of that which should not be. The paper's critical central thesis is finally that the fundamental contradiction of the ethics of compassion is condensed in the figure of the ascetic. The sense of release (*Gelassenheit*) demanded by the negation

of the will excludes the normative reference to the other which Schopenhauer had once indicated as the mainspring of moral action.

*Keywords:* Rousseau, Schopenhauer, the ethics of compassion, ascetic.

HYE YOUNG KIM

### **An Existentialist Analysis of Forgiveness and Gratitude**

The emotions that are dealt with regularly in Existentialism are angst, guilt, fear, concern, and perhaps shame and such. This is because the core of the understanding of human existence is profoundly related to the finiteness of human existence. Furthermore, the influence of the Christian theological interpretation of human existence on Existentialism is undeniable, whether it was fighting or embracing the ideas of the Christian tradition. Either way, the emotions that were highlighted in existential philosophy or other fields of philosophy did not go further beyond the scope of these certain emotions. I attempt to examine other emotions that have been rather neglected in traditional philosophy, such as gratitude and forgiveness, and show how these emotions are fundamentally related to the understanding of our existence.

*Keywords:* existentialism, finiteness, forgiveness, gratitude.

CSABA OLAY

### **Alienation**

The paper discusses two aspects in the work of three paradigmatic thinkers of alienation: Rousseau, Marx, and Lukács. The first issue can be expressed with the question of what should be reappropriated in overcoming alienation? The second point concerns the question of how we experience our being alienated?

With regard to Rousseau, I examine the structure of what might be called a precursor conception of alienation. Alienation in his sense has the structure of possession and subsequent disappropriation of man's original constitution. Taking Marx's more specific concept into account, it can be pointed out that the general structure of alienation might be described with the possession – disappropriation – reappropriation formula. But in Rousseau we have a simplified version of alienation in the form of hypothetical possession – disappropriation.

The second, shorter part of the paper deals with Marx' theory of alienation. It will be shown that normative basis of alienation in the early Marx is the concept of man's self-realization in the working process that takes place in a double movement of a prior objectification and a following re-appropriation. The last part of my argumentation is dedicated to Lukács's theory of reification in *History and Class Consciousness*. I discuss Lukács's critique of capitalist society with an eye on how the concept of reification partly carries on and partly modifies Marx's conception of alienated labor. This part of the paper shows that even Lukács could not clarify how non-alienated conditions should be conceived.

*Keywords:* alienation, reification, Rousseau, Marx, Lukács.

CSABA PLÉH

**Narrative Identity in Its Crises in Modern Literature**

Modern memory schematization started with the schema notion of Frederic Bartlett. Bartlett used stories to support his constructionist theory. New structural approaches to stories have emerged in the work of Colby, Rumelhart, and others. Like Bartlett, they were looking for underlying social schematization and constraints. Narrative patterns promised to provide a substantial anchoring point for the otherwise elusive concept of schemata proposed by Bartlett. In contrasting the alternative models, the ones relying on elementary social attribution molecules were the empirical winners. This consensus affiliates memory schematization with theories that treat elementary sociality as a basic, non-constructed feature of the human mind. Parallel to research on narrative memory, a narrative movement was initiated in psychology by Jerome Bruner.

These psychological narrative theories were extended to the issue of decompositional theories of the Self. These usually approached from two angles: from the body image and from social roles and relations. The later social role based models turned into narrative construction theories of the Self, promoted by Paul Ricœur and Daniel Dennett, coming from very different philosophical heritages. These theories also relate to the issues of how modern novelists and theorists from Milan Kundera to David Lodge proposed novel writing as a factor in the birth of the modern Self. How then do they postulate a reader who considers the action plans, and how did they try to relate them to layers of action, intention, and feeling?

The twofold heritage was also accompanied by a third attitude that proposed the unfolding of the self in autobiographical story telling practices. The search for explanatory principles underlying schemata by the experimentalists, the use of autobiographical narratives, and the cultivation of broken narrative patterns in modern novels can be seen, as Jerome Bruner proposed, as a modern way to present the traditional dualism between *Naturwissenschaften* and *Geisteswissenschaften* as a duality of a categorical and a narrative approach to the human mind.

*Keywords:* schematization, narrative theory, intentionality, novel writing, Self and narrativity.

JUDIT SZALAI

**The 'Reasons of Love' Debate in Analytic Philosophy:  
Reasons, Narratives, and Biology**

This paper is a contribution to the “reasons of love” debate in analytic philosophy. The claims it is meant to substantiate are the following. First, the “reasons-based”/“no-reason” views do not constitute a genuine theoretical dilemma: we do not love persons for either abstract properties that several individuals can share, or for some elusive “ipseity”. Second, descriptive and normative approaches (why persons love and why they *should* love others) should be clearly distinguished. Third, making distinctions between different forms of (romantic or quasi-romantic) love advances matters a great deal,

as reasons apply to these in different ways. The interplay of different factors in loving persons (personal properties of the beloved as reason-giving, joint history, and bio-psychology) are especially relevant to the types of romantic love.

*Keywords:* love, reason, history, narrative, biology.

TIM THORNTON

### **Narrative Identity and Dementia**

Given both that dementia undermines memory and the longstanding view that personal identity depends on memory, it seems likely that dementia undermines personal identity. That connection has, however, been criticised by those philosophers and healthcare professionals who subscribe to a narrative account of identity. While the capacity to author a self-narrative is also threatened by dementia, that need not undermine personal identity providing that the relevant narrative can be *co-constructed* with others. In this paper I set out the danger of any such view, explore its motivations and provide a minimal account of the role of narrative in dementia making use of the Wittgensteinian notion of secondary sense.

*Keywords:* dementia, person, narrative identity, co-authoring, Wittgenstein's secondary sense.

DAVID WEBERMAN

### **What is an Existential Emotion?**

My aim in this paper is to make more precise the idea of an existential emotion. The framework for my analysis follows Heidegger's account in *Being and Time*. In the first part, I will argue that Heidegger's notion of *Befindlichkeit* is essentially about what we call emotions and that emotions come in two types: i) moods and ii) object-specific emotions. I will argue that Heidegger takes both types (correctly) to be intentional, that is, directed at or about something. In the second part, I use the notion of existential in a way that applies to certain emotions, asking which emotions can be existential and what makes them so. Is it only moods that are existential? And, among moods, are there certain of them such as *Angst* that have a special claim to being existential in the sense used here? This will lead to me to the third part of the paper in which I pursue the various ways in which emotions can be seen as existential depending on how and what they disclose. In the end, I present a sort of template for the existential character of emotions and suggest how the notion of existential might be understood to apply to other things besides emotions (such as ideas or literary works).

*Keywords:* emotion, mood, existential, Heidegger.