Contents

Dagmar and John Searle, Karlovy Vary Symposium on Meaning, September 1993.

Introduction: John Searle in Czech Context

In the seventies, when the university teaching and official publication policy in Czechoslovakia admitted only one philosophical orientation, numerous Czech and Slovak linguists intensively studied analytic philosophy of language and worked with its conceptual apparatus. The main interest focused, quite naturally, on the works by John Austin and John Searle. For many philosophy students of that time, the first encounter with the analytic philosophy has been mediated by debates among linguists (in more or less closed circles): the role of the initiating work was then typically played by John Searle's *Speech Acts* and some of his articles later collected in Searle (1979).[1]

Meeting their admired author in person during his repeated visits to the Czech Republic since the early nineties was an exceptionally inspiring (and most pleasing) experience for all these scholars.[2] The visits started with a unique series of lectures in May 1991 held at various Prague philosophical and linguistic departments, with topics ranging from the state of the US philosophical scene through fundamental issues of contemporary philosophical debates (like rule-following or the realism-antirealism dispute) to special questions of joint interest for linguists and philosophers of language (like the status and functions of explicit performatives).[3]

The impact of these lectures and debates with John Searle (together with his philosophical writings) on the formation and development of Czech analytic philosophical thought has been enormous and the interest they have attracted among Czech philosophers and linguists

---

[1]   Among many others, this was the case of one of the editors of this volume. This is perhaps the right opportunity to mention the linguistic seminar held in the early seventies at the Faculty of Philosophy in Brno, focusing on detailed work with John Searle's definitions of speech acts. The seminar was led by Mirek Čejka, probably the most competent Czech interpreter of John Searle's and John Austin's work during the so called "normalization" period.

[2]   We are especially pleased to add that these visits gave us the opportunity to get acquainted also with John Searle's wife Dagmar.

[3]   The text of the lecture 'Is There a Problem About Realism?' appeared in Searle (1992).

continued in the following years.[4] It culminated during John Searle's visits to Prague and Ústí nad Labem in May 2011 and June 2012.[5] The hopelessly overcrowded *Conference Centre* of the Prague's *Institute of Philosophy* has hosted his lecture on *The Intentionality of Visual Perception* (whose transcription appears in this volume) and, one year later, on *Language and Social Ontology.*

Both lectures have been followed by workshops devoted to John Searle's philosophical work. The papers presented during them, or on the Searlean colloquium in Ústí nad Labem, gave rise to the idea of this volume. Several other papers appeared after this plan became commonly known. Honestly speaking, it is hard to imagine that we could collect so many original Czech articles related to the work of any other philosopher of our times. The range of the material presented in this volume illustrates (at least partially and rather selectively) how profound and manifold is the impact of John Searle's work on Czech philosophers, linguists and literary theorists.[6]

It is not difficult to say what the reasons for such an enormous interest are. The first is, obviously, the indisputable significance and exceptionally wide thematic scope of John Searle's work. You can hardly understand (and even less participate in) contemporary debates on meaning, reference, communication (in its "ordinary" forms as well as in rather specific ones, such as the fictional discourse or the highly "ritualized" use of language within the operation of various institutions), on the nature of intentionality, consciousness, artificial intelligence, social reality and its institutions etc., unless you are familiar with John Searle's views on these matters. Second, these views are presented with precision and clarity which can serve as a paradigm of the virtues typically proclaimed (but not always observed) by analytic philosophers. As a consequence, the author's statements, arguments and theoretical constructions are highly transparent and directly open to critical exam-

---

[4]   Most of these visits have been organized by the *Center for Theoretical Study* and initiated by its head Ivan M. Havel. Papers presented at the colloquium on John Searle's philosophy organized by the *Center* appeared, together with others, in Smith (2003).

[5]   Both visits have been organized by Josef Moural from the Philosophical Faculty of J. Ev. Purkyně University in Ústí nad Labem.

[6]   As a special bonus for the reader, we have included (with great pleasure) also a paper by our Budapest colleague Zsofia Zvolenszky, which has been presented at the Prague workshop on John Searle's philosophy in June 2012.

ination; moreover, the concepts and distinctions introduced by him in this manner into a theoretical discourse find immediately an operative application as instruments of philosophical analysis.

Those who appreciate philosophy as a unique opportunity for speaking about unspeakable will not be impressed by these virtues. For analytic philosophers, on the contrary, they should count as a criterion for seriousness of any theoretical work as well as of the author's respect to the reader. Our respect to John Searle and his work gave rise to this volume and the authors did their best to do justice to values they find most convincingly exemplified in his way of doing philosophy. They have not found a better way of expressing their appreciation of the role his philosophical work and his personal engagement played in the constitution and rise of analytic philosophy in the Czech Republic.

We are grateful to John Searle for his kind approval to publish a transcription of one of his Prague lectures and in particular for substantial, detailed and creative reactions to each of the 15 papers collected in this volume. In these responses, the reader receives the most valuable source for better appreciation of John Searle's views on many significant philosophical topics, their background and their consequences.

Our thanks belong also to the *Philosophical Institute of the Czech Academy of Sciences* for generous support of this project and to the editors of *Organon F* for cooperating on its implementation.

<p style="text-align:center">* * *</p>

The volume is opened by a transcription of the lecture on *The Intentionality of Visual Perception*, given by John Searle at the Institute of Philosophy of the Czech Academy of Sciences in Prague on May 23, 2011. We are grateful to Irma Hvorecká for having reconstructed the talk from (technically not quite perfect) video recording and to John Searle for having revised the resulting text for the publication in this volume.

The papers of other contributors are divided into four sections, rather roughly and with numerous thematic overlaps. The first one, focusing on the problems of **ontology**, begins with two attempts to specify and critically evaluate John Searle's position in the realism-antirealism dispute: *Lukáš Zámečník* analyzes the relations of Searle's external realism to the system of commonly shared beliefs and communicative commitments, while *Tomáš Marvan* argues that Searle's position is in fact compatible with ontological constructivism and defends the idea of "privileged conceptual scheme", rejected by Searle. In the third paper

of this section, *Vladimír Havlík* discusses Searle's account of ontological emergence and raises several questions concerning its explicative potential.

A special challenge (in particular for those who feel committed to the naturalistic position) represents the emergence of consciousness and of behaviour based on free will – phenomena that are thoroughly discussed in the first two papers of the second section, devoted to the **philosophy of mind**. *Martin Pokorný* develops an account of consciousness which is (intended to be) in the Searlean line, while *Juraj Hvorecký* discusses Searle's solution of the (apparent) conflict between the universal determinism and the possibility of free will. In other contributions to this section, *Pavla Toráčová* approaches Searle's theory of social institutions as a source of the explanation of the nature of intentionality; *Tomáš Hříbek* addresses the dispute between Searle and Davidson concerning the attribution of thought to prelinguistic animals and suggest a position which should enable to appreciate relevant arguments raised on both sides; *Petr Koťátko* defends and applies the Searlean internalist way of construing the content of thoughts and communicative acts, which (as he attempts to show) reflects and incorporates some aspects of the externalist initiative.

An important part of this discussion (and of the externalist/internalist debate in general) concerns the role of proper names – a topic of the first two papers of the **philosophy of language** section. *Zsofia Zvolenszky* demonstrates how Searle's theory of proper names, in its authentic form, not identifiable with the cluster-of-descriptions doctrine commonly attributed to Searle, is (or can be made) safe against the main objections raised by its critics. *Marek Nagy* suggests how to approach proper names from the perspective set up by Searle's theory of social reality, in particular by exploiting his notion of status function. *Vít Gvoždiak* proposes an interpretation of Searle's achievements in the philosophy of language, as well as of his analysis of the constitution of social reality, from the point of view of semiotics. In the last contribution of this section, *Tomáš Koblížek* addresses some issues of the well-known discussion between John Searle and Jacques Derrida: he defends (against Searle's objections) Derrida's account of the purity of concepts (as independent on the results of their application) and questions Searle's account of the relation between expression meaning and utterance meaning.

The last group of papers deals with the applications of Searle's philosophy of language (in particular of his speech act theory) in the

theory of *fiction* and in the analysis of **metaphors**. *Jiří Koten* discusses the applicability of Searle's account of fictional discourse (as based on pretended assertions) in the theory of film, while *Jan Tlustý* exploits Searle's theory of fiction in analyzing the distinction between fictional and factual autobiography. In the last paper, *Jakub Mácha* confronts Searle's and Davidson's account of metaphors and argues that Searle's theory can meet most of Davidson's objections.

Although this collection has been intended as homage to John Searle and congratulation to his anniversary, the real gift was received by the contributors (and with them all the readers) in John Searle's thorough and attentive replies. They appear in the last section of the volume.

The Editors

References

SEARLE, J. (1969): *Speech Acts. An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

SEARLE, J. (1979): *Expression and Meaning. Studies in the Theory f Speech Acts*. Cambridge: Cambridge University Press.

SEARLE, J. (1983): *Intentionality. An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.

SEARLE, J. (1992): Is There a Problem About Realism? *Filosofický časopis* XL, vol. 3, 413-433 (Introduction by P. Koťátko).

SMITH, B. (ed.) (2003): *John Searle*. Cambridge: Cambridge University Press.

Perceptual Intentionality

John R. Searle

Perception looms large in our philosophical tradition. There are a number of reasons for this, but two stand out. First, for the three centuries after Descartes epistemology, especially skeptical epistemology, was at the center of Western philosophy; and of course there is really no way to face these issues without advancing a theory of perception. A second reason is that philosophy is generally concerned with the relation of human beings and reality, and the primary conscious medium by which humans relate to the real world is the medium of perceptual experience.

Unfortunately, the entire philosophical tradition rests on a disastrous mistake. This mistake in turn rests on a failure to understand the intentionality of perception. So the article will have two aims: I want to expose and correct what I will call the Bad Argument, which afflicts just about all of the major philosophers after Descartes, in fact all known to me; and second I want to provide at least the beginnings of an account of the intentionality of perceptual experience.

1   The Bad Argument

The account of perception that I will present is a form of *direct realism*, according to which, in a typical perceptual experience, we are directly aware of objects and states of affairs around us. "Directly" means that the relationship is unmediated. Looking at the desk in front of me is not like seeing the desk on television or reflected in a mirror. In those cases I do not see the desk directly; right now I see it directly. Traditionally direct realism is opposed to representative realism, according to which there is a real world out there; but we cannot perceive it directly. We only perceive representations of it in the form of our impressions, ideas, or sense data. Famous representative realists are Descartes and Locke.

Because direct realism is so obviously consistent with both our experience and with common sense, Why has it been so frequently denied? Here is the amazing fact. None of the Great Philosophers in the

tradition that I am familiar with – and by Great Philosophers, I mean Descartes, Locke, Berkeley, Hume, Leibniz, Spinoza, Kant, and I would have to add Mill and Hegel – were direct realists. They all, in one way or another, accepted an argument that goes back at least to the 17th century, and for all I know it may even go all the way back to the Greeks. The standard argument against direct realism is called the Argument from Illusion. And here is how it goes: I said that I see the desk in front of me, but suppose I were having a hallucination. I could be having an experience exactly like this, completely indiscriminable from this one with exactly the same content as this one, and be having a total hallucination. You could have the same experience without there being a desk on the other end of the experience. But now comes a crucial step. The character of the experience in the hallucination case and the character in the veridical case, the bad case and the good case respectively, is the same; so any analysis of one has to be applied to the other. But in the bad case, though one is not aware of a desk, one is certainly *aware of* something. One is conscious of something, and, at least, in some sense, one "sees" something. That something cannot be a material object, because there is no material object there in the case of the hallucination. Yet I am aware of something. Let us give a name for such somethings; they were called "ideas" by Berkeley, Descartes, and Locke, "impressions" by Hume, and came to be called "sense data" in the 20th century. So I will stick with "sense data" and say we are aware of a sense datum. A sense datum is an entirely mind dependent, ontologically subjective entity. Now, by the principle that both the good case and the bad case should receive the same analysis, it follows that in the good case I am not aware of a material object but only of sense data. But now it seems to follow that in all experiences I am aware only of sense data, not of mind independent material objects. And the question then arises, What is the relationships between the sense data that I do see and the material object that apparently I do not see? This argument in various forms survives right to the present day. What is wrong with the argument? On the surface, at least, it rests on a simple fallacy of ambiguity. The expression "aware of", "conscious of", and even "sees" in this argument, are ambiguous. The ambiguity can be illustrated by using a very simple and unproblematic example. If I push my hand very hard against the top of the desk, I am aware of the surface of the desk. This is the intentionality sense of "aware of" that has the desk as the intentional object. But it is obvious that I am also aware of a painful sensation in my hand

assuming I push hard enough. So which am I aware of: the painful sensation in my hand or the top of the desk? After all there is only one object there and one experience there. Which one is the genuine case of awareness? The answer is that the expression of "aware of" is used in two different senses. In the intentionality sense the desk is the object of the awareness, and I am aware of the desk. But there is another sense being exhibited here, where I am "aware of" a painful sensation; and that case is not one of intentionality, because the awareness and the painful sensation are identical. This is a crucial point. Where intentionality is concerned the sensation is not identical with the object, but there is another constitutive sense in which the awareness and the thing one is aware of are identical. The proof then that there two different senses of "aware of" are being used here is that the semantics are different. In the intentionality sense, the subject S has an awareness A of object O implies A is not identical with O; but in the constitutive sense where subject S has an awareness of entity O, A is identical with O, the painful sensation and the awareness are identical. Now let us apply this to the famous Argument from Illusion that we considered earlier. In the sense in which I am aware of an object when I look at the desk, the intentionality sense, in that sense when I have a hallucination I am not aware of anything. There is nothing there; hence I could not be aware of anything. Nonetheless, I am having a conscious visual experience and it is tempting, given the way our language works, to erect a noun phrase to stand for that awareness and make it into the object of the verbs of perception. So "aware of", "conscious of", are used in two different senses. We feel immediately hesitant to say that one "sees" anything in the hallucination case, so we are tempted to put sneer quotes around "sees". But what is going on, I hope, is obvious and clear. In every case there is an ambiguity in the crucial phrases "aware of" or "conscious of"; because in the intentionality sense in which I am aware of something when I see it, in the case of the hallucination I am not aware of anything. I have a conscious experience, but that conscious experience is not itself the object of the experience; it is identical with the experience.

Once pointed out that this is such an obvious fallacy it is hard to see how anybody could have made it; but nonetheless there it is and it produced the idea that is common to the Great Philosophers that one does not perceive the world or does not perceive it directly. One perceives only the contents of one's own mind, one's own sense data. Ironically, the argument is repeated by people who think that they are defending

naïve or direct realism. They are called "disjunctivists", and they accept the validity of the argument but wish to reject its conclusions on the ground that they think the first premise is false. The first premise says the hallucination case and the veridical case have exactly the same content. The experience is indistinguishable for the obvious reason that it has exactly the same content and the same form of awareness. Disjunctivists think that in order to protect naïve realism you have to reject that premise. Thus in order to save naïve realism they reject an obviously true premise. And what I am suggesting here is that you can have naïve realism and still avoid making the fallacy that led to its rejection. To repeat this point: the disjunctivists correctly see that the conclusion of the argument – the denial of direct realism – is false; but they think in order to save naïve realism they have to reject the true premise that the content of the good and the bad cases can be exactly the same. They accept the validity of the argument but think it is unsound, because the first premise is false. On the other hand, my argument is that the first premise is entirely true and the argument is *invalid*, because it rests on a fallacy of ambiguity.

Prague Lecture on The Intentionality of Visual Perception, Institute of Philosophy, May 2011

The same fallacy, by the way, afflicts the so-called Argument from Science that says science has shown we can never see objects and states of affairs in the world, but only our own experiences of those objects.

The story goes that the neurobiological events that lead from the sensory receptors to the conscious experience determine that all that we can ever be aware of is the conscious experience. I hope it is obvious that this is the same fallacy: the fact that one can give a causal account of how the conscious experience occurs does not show that one does not see the objects and states of affairs on the other end of the conscious experience. To suppose that is to suppose that the experience itself is the object of perception. And that is the Bad Argument all over again. I am not going to go through the entire history of philosophy to show how the Bad Argument keeps creeping up in Descartes, Locke, Berkeley, Hume, etc. I do not know that Kant ever uses the argument in this form, but he certainly accepted the conclusion.

2   The Intentionality of Perceptual Experience

I am going to assume that the Bad Argument has been refuted and that we can accept direct realism as the accurate account of veridical perception. How does it work? Before answering the question I want to situate the discussion within contemporary and recent philosophy. Traditionally, analytic philosophy is obsessed with truth conditions. To analyze a concept or a sentence is to give its truth conditions. Thus, to take two famous examples, Frege makes his distinction between *Sinn* and *Bedeutung* in his analysis of the truth conditions of identity statements, such as "The Evening Star is identical with the Morning Star". Again, Russell in his Theory of Descriptions tried to analyze the truth conditions of sentences, such as "the King of France is bald", which contain an apparent reference to an entity that does not exist. In my book *Intentionality* (1983) I extended this project from sentences to mental states, and I attempt to analyze not just truth conditions but what I call *conditions of satisfaction*, generally; so that I would be able to give the form of the conditions of satisfaction not just of sentences but of beliefs, desires, hopes, intentions, and perceptual experiences. Typically, philosophers do not worry about how the sentence or expression gets its conditions of satisfaction; they just want to know what are the conditions of satisfaction. And the answer, in general, is that in the case of linguistic elements the conditions of satisfaction are imposed by convention, and the form of that imposition is called assigning a meaning to the sentence. The meaning of the sentence is that which determines the conditions of satisfaction, and the sentence has its meaning by the

conventions of the language of which it is a part. Our present task is much more difficult than that. Perceptual experiences with their raw phenomenology do not in general have conventions that fix their conditions of satisfaction. All the same they do determine conditions of satisfaction. Our question is, How does it work?
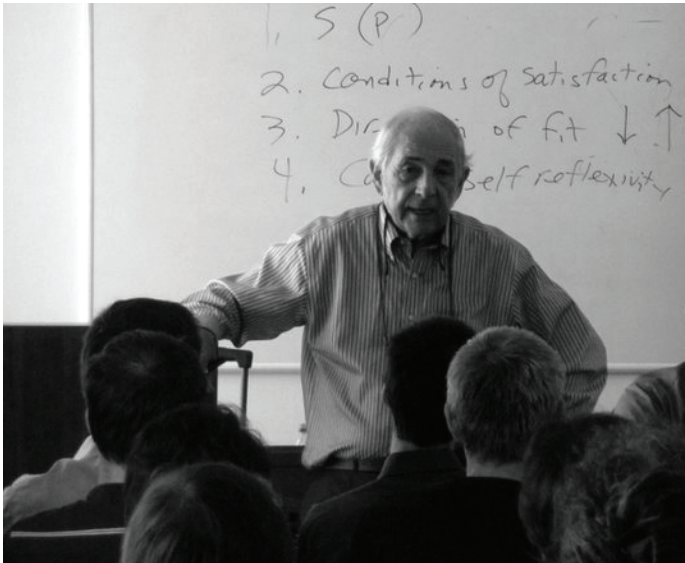
Prague Lecture on The Intentionality of Visual Perception, Institute of Philosophy, May 2011

The key to understanding perception is to recognize both that it has intentionality in the philosopher's sense, according to which intentionality is simply the directedness or the aboutness of the mind, but it has a special form of intentionality that I call *presentational intentionality*. Perceptual experiences, visual or otherwise, are in the intentionality sense, directed. But the intentionality of visual experience differs from the philosopher's favorites of belief and desire in a number of important respects. Beliefs and desires are typically representations of objects and states of affairs in the world. But when I see something I do not just have a representation; I have a direct presentation of the object. My visual experience, so to speak, reaches right up to the object.

What does it mean exactly to say that the perceptual experience is a presentation and not just a representation? There are a number of features of perceptual presentations. I do not have the space to go into all of them, but I will mention some of the most salient. The most imme-

diate and obvious feature of perceptual presentations is that they are directly caused by the object itself. The present features of the object I am seeing are the conditions of satisfaction – what makes the visual experience "veridical" – but they are experienced immediately, and they are experienced as causing the very perceptual experience that has the object and the features as the rest of the conditions of satisfaction. So there is a causal self-reflexivity to perceptual experience in that it is only satisfied if the object or state of affair seen causes the experience itself that has those conditions of satisfaction. Memories and prior intentions also have this causally self-reflexive feature, but they are not direct presentations. You are not immediately aware of the object you *remember* in way that you are if you actually *see* it.

Prague Lecture on The Intentionality of Visual Perception, Institute of Philosophy, May 2011

A second feature of the presentational intentionality of visual experience is that it is always of the here and now. Everything you see, touch, hear, you are hearing as existing right here and now. This, as a feature of the phenomenology, remains true even in cases where we know the object is not right here and now. If I look at the star that I know ceased to exist millions of years ago, all the same, I am seeing it as if it were visible right here and now. The third feature of perceptual experiences that differs from features of other more commonly discussed forms of intentionality, such as beliefs, is that they are experienced as non-detachable from the conditions of satisfaction. If I close my eyes and think about objects, I experience certain representations of these objects and I can shuffle them around at will. But if I open my eyes and

look directly at objects in front of me, I cannot shuffle my experiences around at will. They are experienced as directly connected to the object that I am seeing, and this is a feature of the phenomenology that holds even for hallucinations. Even if it is a hallucination and even if I know it is a hallucination, all the same, I do experience it as non-detachable from the object that it is as if I were seeing.

So presentational intentionality differs from such things as beliefs and desires in at least these three crucial respects. It is causal and it is experienced as directly causal, that is to say, it is experienced as directly caused by the object or the state of affairs perceived. Secondly it is always indexical; its conditions of satisfaction are always immediately tied indexically to the occurrence of the experience itself. Every time I see something I see something as existing right here and now, and every time I see an event I see that event as occurring right here and now. Third, it is because of these two features that it is not detachable from the conditions of satisfaction. I cannot shuffle it around as I can with other sorts of mental representations.

But these reflections so far do not answer our question, How does it work? And I now need to explore that question in some detail.

3   The Subjective Visual Field and the Objective Visual Field

If you close your eyes and put your hand over your eyes, you will have an experience which is something like seeing a black area with yellow patches in it. You can try this for yourself. Now, strictly speaking you do not see anything, because your eyes are closed; but you do have conscious experiences which you would naturally describe in the vocabulary that I just used. You do not see anything, because you cannot see anything with your eyes closed. But, nonetheless, you do continue to have conscious visual experiences. This area of your conscious experiences I propose to call the *subjective visual field*. Now open your eyes and the subjective visual field is suddenly filled with everything that constitutes your conscious awareness of the objects and states of affairs in your vicinity. To have a name for what you can perceive I want introduce another technical term, the *objective visual field*, which consists in all of the objects and states of affairs that are visible from your point of view and given your physical condition at any particular time. The objective field is ontologically objective in that it is perceivable by any similarly endowed and similarly situated person. The chair that I see

is part of the objective visual field and it is ontologically objective, because it is accessible to anyone equally. But, in addition, we have found it necessary to postulate the subjective visual field, which consists entirely of the conscious experiences going on in my head. Much of the rest of this article will be about the relations between the subjective visual field and the objective visual field.

The most important thing, I can tell you immediately, is that in the objective visual field everything is seen or at least can be seen; in the subjective visual field nothing is seen nor can be seen. This is not because the entities in the subjective field are invisible, but because they are the seeing of anything. The idea that the entities in the subjective visual field are themselves seen is the basis of the Bad Argument. The awareness itself is erected into the object of awareness; and this denies the intentionality of perception, because the experience is not the intentional object of the experience, it is the experience itself.

The question I want to address now is, How does the raw phenomenology of the visual experience, the actual qualitatively subjective entities in my head, set the conditions of satisfaction – that is how is it that they have the content that they do? Another way to put this question is to ask, What fact about my subjective experiences makes it the case that I am seeing or at least seem to be seeing specific types of objects and states of affairs in the objective visual field?

Prague Lecture on The Intentionality of Visual Perception, Institute of Philosophy, May 2011

This is not a trivial question and I will not fully answer it in the scope of this article, but I can at least suggest some general principles which have to govern any answer.

Two traditional answers to this question both fail. They are resemblance and causation. According to the resemblance theory the visual experience presents red objects because it is itself red. There is a resemblance between the visual experience and the object perceived. Everything is wrong with this answer. To begin with, the visual experience could not possibly itself be colored; and it is not itself perceivable. Causation will not do as an answer either. According to this answer the experience of red has red as its conditions of satisfaction, because it is caused by red things. This will not do, because causation by itself does not carry any intentional content. Suppose that every time I see a red object it causes in me a pain. This does not have the consequence that "red" means pain. Neither resemblance nor causation by themselves are the answer to our question. What is the answer to the question?

4 The Hierarchy of Perception and Basic Perceptual Features

Perception is hierarchical structured in that the top level depends on our ability to see the bottom level until finally you get down to the basic level of things that you can see without seeing anything else by which you can see them. So, take an example, I do not just see colors and shapes, but I see cars, trees, houses; and indeed I do not just see any cars, trees, and houses, I see *my car* in the parking lot, or *my house*. The point about the hierarchical structure is this: in order to see that it is *my car*, I have to see that it is a car of a certain make and age; but in order to see that it is a car of a certain make and age, I have to see that it is a car at all; and in order to see that it is a car at all, I have to see that it has certain color and shape. Color and shape are basic relative to the other features, so being *my car* is less basic than being a car of a certain type; being a car of a certa in type is less basic than being a car; and you go down until you finally reach basic perceptual features, such as color, shape, movement, etc. The intuitive idea is that a basic perceptual feature is one you can see without seeing something else by way of which you see it. This cannot be quite the right way to put it, because on this criterion we cannot distinguish color and shape; and maybe the right way to describe the basic feature is to say "colored shapes" or "shaped colors". The intuitive idea is clear: perception is hierarchically structured and hierarchy bottoms out in basic perceptual features; and corresponding to the hierarchy in the objective ontology is a hierarchy

in the subjective visual field. The experience of the higher level features – such as being my car – requires the experience of the bottom level features – such as color and shape.

So far then we have both the notion of basic perceptual features, which are ontologically objective features of the world that are perceivable, and the notion of basic perceptual experiences that have the basic perceptual features as their objects, as their conditions of satisfaction. And we can now narrow our question so that it is much more manageable. How does the raw phenomenology of the basic perceptual experience fix the basic perceptual features as the objects of perception, as their conditions of satisfaction?

A condition of adequacy on our answer to this question is that the connection between the raw phenomenology and the conditions of satisfaction must be internal or essential. That is, for example, it could not be this very experience if it was not an experience of seeing, or at least seeming to see, something red. Now what fact about the experience gives it that intentional content? And remember in answering that neither resemblance nor causation by themselves are going to be enough.

To answer the question I propose to emphasize the fact that our perceptual relations to the external world are causal and experienced as causal throughout. The whole Humean tradition about causation makes it difficult to see that causation is everywhere and is experienced as everywhere. Hume taught us we cannot experience necessary connection. I think we experience it pretty much all day long. Wherever we consciously perceive or engage in action we experience objects causing perceptual experiences in us and ourselves causing bodily movements and other sorts of changes in the world. Causation, to repeat, is everywhere. We live in a sea of causal relations and we constantly experience causal connections. The reason that this experience has red as its conditions of satisfaction is because it is the essence of red, it is part of the very definition of "red", that it consists, at least in part, in the ability to cause this sort of experience. It is tempting to put this point conceptually, but that would be misleading if it gave us the idea that in order for an animal to be able to perceive red it has to have some conceptual skills. The point rather is this: it is of the very essence of something being red, and hence of the concept of red, that red consists, at least in part, in the ability to cause this sort of experience. What I have said about "red" is, of course, true of all other colors. Their essential feature is this ability to cause these sorts of experiences.
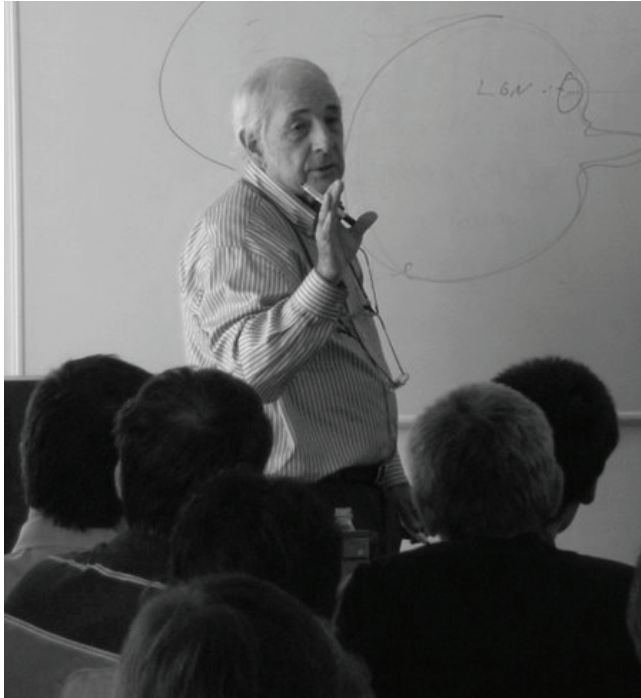
Prague Lecture on The Intentionality of Visual Perception, Institute of Philosophy, May 2011

The analysis is a bit trickier for the so-called primary qualities, but I think it works for them – or at least for the two dimensional ones – as well. Lines and shapes are in part defined in terms of their ability to cause such and such experiences. So a straight line causes this sort of experience. Notice it will not do to say that a straight line is one that "looks like this", or red is a color that "looks red" under these conditions; because "looks like" and "looks red" have a meaning deriving from "is". To look red means to look *to be* red, looks *as if it is* red; and that is what we are trying to explain. The fact about it that makes it look as if it is red is what needs explaining, and the explanation I am offering is that red, by its essence, consists in the ability to cause this sort of experience.

Well, this does not give us an analysis of a very rich array of perceptual features. We are down just to the bare bones of basic features. What about the others? I do not have the space to go into detail here.

I am writing a whole book about this. But, just to mention a few of the points: depth or three-dimensionality is not a basic perceptual feature, rather we experience depth because our experience of the basic perceptual features impacts on our nervous system in a way that our Background mastery of the principles of perspective enables us to treat them as carrying three-dimensionality as part of the conditions of satisfaction. As part of the basic experience I see two parallel lines that are getting closer toward the top, but of course I do not just see two parallel lines; I see two railroad tracks receding away from me in the distance. How do I get from the perception of the basic perceptual features to the perception of the distance as the railway tracks move away from me in the distance? The answer is that my Background skills contain a mastery of the principles of perspective, and this enables me to treat the basic two-dimensional features three-dimensionally.

Notice I am not saying we make an inference. We do not need to. We just see distance immediately. I can literally see that the desk is further away from me than the chair. But in order to do that the nervous system must be able to operate on the basic perceptual experiences in a way that imposes the three-dimensional understanding on a visual array in my subjective visual field that can be produced by a two-dimensional stimulus.

What about more complex cases? Suppose I see that this is a California Coastal Redwood Tree, or that the car in the parking lot is my car. Some complex visual phenomena can be defined in terms of the basic features. To learn to identify the Coastal Redwood Tree as a Coastal Redwood is to be able to identify it in terms of its basic features: color, shape, texture, etc. Being a California Coastal Redwood Tree is defined as having a certain combination of basic perceptual features – at least as far as visual experience is concerned. Of course scientists will give a fancy scientific definition in terms of the DNA, but what we are talking now is what I can literally see. The phenomenology of the complex is made out of the phenomenology of the simple.

We also need to explain recognition of previously experienced objects. So, for example, my perception of my car carries much more than just complex perceptual features. The facts about it that make it the case that I perceive it as mine involve the notion of *recognition.* And how does recognition get into the phenomenology? In the case of seeing it as my car, I do not just see it as a car having such and such features; but I see it as identical with a car that I have experienced earlier on count-

less occasions of having such and such features. That is, as far as the phenomenology is concerned, this perceptual experience is simply the latest in a sequence of perceptual experiences all of which are caused by the perception of my car.

5   Conclusion

This article is only really the beginning of an account of how perceptual experiences fix conditions of satisfaction, but I hope the direction is clear. There are three points that I wish to emphasize. One, direct realism is preserved once we abandon the Bad Argument and resist any temptation to resurrect it. Two, perception is obviously intentional; but it has a special type of intentionality that I call presentational, rather than just representational. And third, we can begin to explore how the raw phenomenology of the perceptual experience sets the condition of satisfaction that it does. For an animal for perceive a basic property as F is to for it to have a perceptual experience that it experiences as caused by something whose essential character is the ability to cause F. I am not, of course, saying that animals think all of this, nor do humans except for rare philosophers; but this is what is going on in the phenomenology of their experience, and indeed if you watch an animal engaged in complex conscious behavior – a dog digging for a bone, for example, or chasing a cat – you see precisely this combination of the conscious experience of causation and the identification of intentional objects.

Department of Philosophy
University of California
314 Moses Hall 2390
Berkeley, CA 94720-2390
USA
searle@berkeley.edu

References

SEARLE, J. R. (1983): *Intentionality: An Essay in the Philosophy of Mind.* Cambridge: Cambridge University Press.

I. Ontology

External Realism as a Non-Epistemic Thesis[1]

Lukáš Zámečník

Palacký University, Olomouc

Abstract: In this paper I concentrate on evaluating Searle's concept of external realism as a non-epistemic thesis. Above all, I evaluate the role it plays in our system of beliefs, knowledge and communicative obligations. There are two important questions: (1) whether a return of transcendental philosophy creates a positive alternative to philosophical naturalism (Quine); (2) whether for the same purpose Fraassens' constructive empiricism suffices.

Keywords: condition of intelligibility, constructive empiricism, external realism, point-of-view invariance, naturalism.

> *So, where does point-of-view invariance come from? It comes simply from the apparent existence of an objective reality – independent of its detailed structure. Indeed, the success of point-of-view invariance can be said to provide evidence for the existence of an objective reality. Our dreams are not point-of-view invariant. If the Universe were all in our heads, our models would not be point-of-view invariant.* (Stenger 2006, 187)

In recent decades we have witnessed the return of metaphysics to the analytical tradition. John Searle is also sensitively emancipating and elaborating several metaphysical concepts without, however, falling into the familiar pitfalls of traditional metaphysical controversies. In this paper I concentrate on evaluating Searle's concept of external realism. Above all, I evaluate the role it plays in our system of beliefs,

---

knowledge and communicative obligations. I also detract from some of Searle's conclusions presented at the end of the eighth chapter of his study 'The Construction of Social Reality' (1997). Whether a return of transcendental philosophy creates a positive alternative to philosophical naturalism (Quine), that is in terms of any fruitful cooperation between philosophers and scientists, is an important question.

Searle defines external realism (ER) as an ontological theory as follows:

> Realism is the view that there is a way that **things** are that is **logically independent** of all human representations. Realism does **not say how** things are but **only that** there is a way that they are. (Searle 1997, 155)[2]

He notes that ER is neither a theory of truth nor the theory of language and not even a theory of knowledge. This clearly and persistently states the ER status as a non-epistemic thesis, and thus rejects the traditional critique of realism in the epistemological tradition of modern philosophy. A certain ambiguity lies in the reasons to link the epistemic approach with antirealism. I believe that this approach could be free of creating ontological theories, this approach does not prefer antirealism to realism, committing both to the pyre because they contribute nothing to our knowledge.

In the previous definition it is important to note the highlighted terms. First, Searle talks about *things*, but notes that it would be preferable to use the neutral designation *it* (as in the sentence "It is raining" does not denote *it* as a reference). Similarly, our ER does not say *how* things are, but *that* they are in a certain (incomprehensible?) way.[3]

This is an important finding, because although Searle repeatedly and firmly states that he does not care about the nature of reality, but about its mere existence, in some situations it is evident that these lim-

---

2   Given the focus only on the 7th and 8th chapters of Searle (1997) in the following text I will include bibliographic information only in direct quotations of this work.

3   In accordance with the noted return of transcendental philosophy it is worth recalling that even the early Wittgenstein can be read through a transcendental lens. Science tells us *how* the world is, but *that* it is remains mystical. "Nicht *wie* die Welt ist ist das Mystische, sondern *dass* sie ist" (Wittgenstein 1969, 148). But this is also an undeniable horizon of our questioning "How?"

its of ER are too restrictive for him. Searle's discussion of the 'Ding an Sich Argument' shows us that, even if the ER provides only a vague *it* (an inconceivable something), it does not follow that knowledge does not concern reality.

> The thesis that there is reality independent of our representations identifies not how things are in fact, but rather identifies a space of possibilities. (Searle 1997, 182)

Second, things are according to the ER logically (not causally) independent of human representations, which brings us to the understanding that the ER cannot be constructed as an empirical thesis, rather:

> ER is thus not a thesis nor an hypothesis but the **condition of having certain sort of theses** or hypotheses. (Searle 1997, 178)

This demonstrates Searle's major conclusion: ER is the condition of the intelligibility of our beliefs.

Whereas philosophers such as Nelson Goodman and Hilary Putnam derive from Searle's recognized conceptual relativity refutation of ER, Searle constructs his thesis that reality may be subjected (adequately) to various alternative (and arbitrary) descriptions. Similarly, Ronald Giere also expresses this in his Perspective Realism (see Giere 2006) with the difference that he does not hold a transcendental position, but notes that the world can be objectively/independently described as it is, which demonstrates the sequence of increasingly sophisticated scientific theories. The transcendental position can be contrarily seen in the expression of the physicist Victor Stenger, who derives his realism from the point-of-view invariance of physical models.[4]

The most valuable aspect of Searle's conception is viewing ER as the basic condition of the intelligibility of our beliefs and differentiates this condition from (mere) truth-condition. At this point it becomes evident that the only argument for ER may be a transcendental one:

> (…) the only argument we could give for ER would be a "transcendental" argument in one of Kant's many senses of that term: We assume that a certain condition holds, and then try to show the presupposition of that condition. (Searle 1997, 183)

---

[4]   "The models of physics cannot depend on any particular point of view" (Stenger 2006, 57).

ER is thus the background presupposition on normal understanding, unlike the others BP, is it fundamentally pervasive and substantial.

Searle uses ER to reject phenomenalist idealism:

> What we should say is this: A public language presupposes a public world in the sense that many (not all) utterances of a public language purport to make references to phenomena that are ontologically objective, and they ascribe such and such feature to these phenomena. (Searle 1997, 186)

This is the most important finding of Searle's position and it introduces the definition of a transcendental dimension of the contemporary analytic tradition. The *World*, of which Searle speaks, so as to grasp the most appropriate equivalent of the *Kantian regulatory idea* of the world, also acts as *Davidson's world*, which is an indispensable element of the triangulation. The world that Davidson and Searle posit is the transcendental condition of the intelligibility of our beliefs. ER is not a condition of knowledge, but an intelligible condition. This conclusion should resonate for a long time because it is the position from which it is possible to criticize not only antirealists of any genre, but also an escalation alternative of the naturalistic tradition of analytical philosophy, which does not admit any non-empirical thesis.

Evaluating Searle's conclusions, we note that Searle does not prove the truth of ER, but only its transcendental sense, and consequently he concedes that he did not refute solipsism. I think it shows us the possibility of an alternative approach. One can renounce discussions of realism, antirealism and solipsism and be content with a pragmatic empirical approach, where the Constructive Empiricism of Fraassens' type seems to be the most adequate.[5] Sober empiricists will not enter into claims about the nature of reality, and will not degrade it to the sum of our representations, leaving only the statement that our representations are constitutive for our descriptions of the world. The regulatory idea of the world provides no clue to our knowledge of the world.

Searle concedes that normal understanding is revisable, but denies that this refutes ER. However, he also states that throughout the history of science there are seemingly ontologically objective phenomena

_____

[5] Owing to Searle's point that normal understanding is not the only understanding that is available, and refers to the problems of quantum physics; see Fraassen (1998).

actually replaced by the objective. The question then arises, where do we derive the justification of their objectivity if the regulatory idea of a world cannot provide such a guarantee? We then revert to the basic problem, because we cannot guarantee replacement of normal understanding only after a strict limit, for this does not imbue the regulatory idea with sufficient potency. Finally we arrive at nothing and can stick to constructive empiricism (not at antirealism).

Despite the convincing accuracy of Searle's conception of ER, we can ask whether transcendentalism is a fruitful variant of analytic tradition. Does it represent a more attractive alternative analytical tradition than a naturalized epistemology? Does it provide us a new understanding beyond the previous tradition?

I am convinced that if, for example, we reproach Quine that his naturalized epistemology could be interpreted idealistically, he would be appalled at what path the analytic tradition has taken. His philosophy has no place for concepts such as realism and idealism.

The corollary of Fraassen's and Searle's view is ultimately the same. Both are convinced of the objectivity of our (scientific) knowledge. But what prevents us interpreting the world in different ways? Is it the regulatory idea of the world?

Let us conduct a thought experiment: Luke is a scientist exploring deep space. Philosophically schooled he knows that the ER is a condition of intelligibility. On his travels he meets several entities that he is able to classify in a developed network of scientific models. Nowhere does he find a hint of other conflicting scientific network models, after examining the entire universe (universe's network) provides a complete theory of reality. One of the entities that Luke classified was SAMAN who has a totally other representations.[6] Basically the only thing which SAMAN matches (unreflectedly) with Luke is the ER. Like Luke SAMAN also classifies everything into totally different scientific network models, without reflecting on alternative networks. Luke and SAMAN cross the frontier into a land of a new kind of "solipsism". Can we argue that in some sense they still share the same world?

The basic motivation for Searle's defense of realism appears to be the need to confront the consequences of antirealistic position that

---

[6] Alternatively, they may have different pure a priori forms, see Kant (1956, 92-93).

erodes the foundations of rationality, and ingresses the irrationality of such thought "systems" as post-structuralism.

I think that we need to combat irrationalism and its attendant anti-realism. However, given that we have no obligation to instill the belief in realism into anyone,[7] it is questionable whether for the same purpose Fraassens' approach suffices.

Department of General Linguistics
Faculty of Arts
Palacký University
Křížkovského 10
771 47 Olomouc
Czech Republic
Lukas.Zamecnik@seznam.cz

References

DAVIDSON, D. (2001): S*ubjective, Intersubjective, Objective.* Oxford: Clarendon Press.

FRAASSEN, B. C. van (1989): *Laws and Symmetry.* Oxford: Oxford University Press.

FRAASSEN, B. C. van (1998): Arguments Concerning Scientific Realism. In: Curd, M. – Cover, J. A. (eds.): *Philosophy of Science: The Central Issues.* New York: W. W. Norton & Company.

GIERE, R. N. (2006): *Scientific Perspectivism.* Chicago: The University of Chicago Press.

GOODMAN, N. (1978): *Ways of Worldmaking.* Hassocks: Harvester Press.

KANT, I. (1956): *Kritik der reinen Vernunft.* Hamburg: Felix Meiner Verlag.

PUTNAM, H. (1988): *Representation and Reality.* Cambridge: MIT Press.

QUINE, W. V. O. (1960): *Word and Object.* Cambridge: MIT Press.

SEARLE, J. R. (1997): *The Construction of Social Reality.* New York: The Free Press.

SEARLE, J. R. (2010): *Making the Social World.* Oxford: Oxford University Press.

STENGER, V. J. (2006): *The Comprehensible Cosmos*. New York: Prometheus Books.

WITTGENSTEIN, L. (1969): *Tractatus Logico-Philosophicus*. Bungay: The Chaucer Press.

---

[7] "Pretheoretically we take external realism for granted, and for that reason it need not be a belief, but is prior to having beliefs" *(*Searle 1997, 195).

Searle on External Realism
and
"Privileged Conceptual Scheme"

Tomáš Marvan

The Academy of Sciences of the Czech Republic, Prague

Abstract: In the paper, I question some of the claims professor Searle makes about external realism, a position I accept. I briefly sum up Searle's position, and then proceed to the mentioned critique. In particular, the target of my paper is Searle's claim that external realist is to shun commitment to any particular ontology. I also point out that Searle's external realism is in some respects difficult to disentangle from ontological constructivism, a position incompatible with external realism. The paper concludes with an apology for the idea of a "Privileged Conceptual Scheme" that Searle views as misguided.

Keywords: John Searle, external realism, constructivism, conceptual relativity, Privilege Conceptual Scheme.

This is an attempt at constructive criticism. Constructive because I subscribe to almost everything Professor Searle has written on realism. Criticism because I do not entirely accept a couple of his claims on the subject and do believe they are in need of a slight revision. Searle is right that a large part of recent literature on realism is vague and confused and that some serious philosophical housekeeping is in order. My aim is to help with the housekeeping in an unambitious, but, I hope, not entirely irrelevant way. I will, first, briefly summarize Searle's realistic position, and then proceed to the mentioned revision.

1 External realism

Searle unswervingly advocates the thesis of "external" realism (henceforward ER).[1] In a nutshell, the thesis consists in the claim that there is a reality totally independent of our representations – words, beliefs, perceptions, pictures, maps, etc. There are objects, features, facts and states of affairs that are logically independent of our representations: even if we and all our representations ceased to exist, a large part of what there is would continue to exist unaffected. Another important feature of Searle's ER is the denial that realism is a doctrine committed to any *particular* ontology. External realist is not committed to the existence of any particular kinds of objects, properties, etc., since we "could be mistaken about how the world is in every detail and realism could still be true" (Searle 1995, 155). Searle further postulates that a properly formulated thesis of "conceptual relativism" is acceptable and fully consistent with external realism. In its proper formulation, the thesis just states that we can devise an indefinite number of conceptual schemes for representing what amounts to a single fact or state of affairs. We can measure weight in pounds or kilograms and it makes no sense to claim that only one of these descriptions correctly captures the properties of physical objects. There is no "Privileged Conceptual Scheme" (Searle 1995, 164). Furthermore, each and every one of our representations is always a part of some system of representations (e. g., a conceptual scheme), and since systems of representations are our creations, a degree of convention or arbitrariness always pertains to the actual form the system of representations takes. This convention or arbitrariness is, though, just a feature of our ways of conceiving external reality and does not affect it. Our systems of representations are constructed by us, but what they represent is, in most cases, no construction of ours. Also, the realist according to Searle need not commit herself to the idea of unconceptualized or, more generally, a-representational access to reality. The fact that we use concepts and other representational devices of our own making in no way implies that we are never in touch with external reality.

All representation occurs within a set of representations and within some representational system. Hence, any representation of the rela-

---

[1] I am going to draw on his 1991 Prague lecture 'Is there a problem about realism?' (later published as Searle 1992) and the seventh chapter of his (1995) book on social reality.

tion between the set of representational states and the representational system, on the one hand and the reality represented, on the other, also occurs within some representational system. But so what? It simply does not follow from the fact that all cognition is within a cognitive system that no cognition is ever directly of a reality that exists independently of all cognition (cf. Searle 1995, 174f.).

No "God's Eye View" on reality is presupposed by the advocates of ER. Finally, ER should be disentangled from any particular theory of truth. Any specific theory of truth, such as correspondence theory, needs a separate defense.[2]

2   A particular ontology

I'll start my critical discussion of Searle's ER with the claim that realists aren't committed to any particular ontology – that their thesis is just a general commitment to there being *a way things are* that is logically independent of all human representations. Searle's position contrasts in this respect with the view of another noted realist, Michael Devitt, who defines external realism in the following way: "Tokens of most current common-sense and scientific physical types objectively exist independently of the mental" (Devitt 1997, 23). To Devitt, Searle's definition of ER would be too minimalist. A *scientific* realist, for example, doesn't just hold that some completely unspecified *something* exists independently of the mental. He holds that *quarks, leptons, mesons*, etc. exist independently of all our representations of them. To say that some unspecified *X* exists independently of representations is uninformative to the point of being completely vacuous. Realism, as I conceive it, is a claim that the world has a certain *structure* which is independent of our representations of it. In other words, the world apart from our representations is not just a totally shapeless blob, a single superentity in which no natural boundaries ("joints") are to be found. Needless to say, we might be more or less wrong about this structure – about the kinds of objects the universe contains, about their features and relations, laws governing their behavior, etc. But pending a very thoroughgoing scepticism, we have a solid evidence that we are tracking at least some bits of this independent structure. Why not say, then, that realism is a claim

_____

[2]   Such a defense is provided by Searle in the last chapter of his (1995).

that quarks and leptons, trees and rivers exist independently of our representations of them?

Searle would not accept this suggestion, for he presents ER as a completely general claim based on modal considerations. In view of the extreme but logically possible scenario that we get *no* feature of external reality right, we must be, he holds, content with the general thesis: there is *something* out there and it is independent of our representations. Thus, even if the world was, say, completely empty, i. e., contained no physical matter whatsoever, ER would, Searle points out, *still be true*: the external emptiness would still be independent of our representations of it. I grant that. But realism as many people, me including, conceive it, is a thesis about the *actual* world, the only one there really is. *As it happens*, the world does objectively contain a vast number of distinct natural kinds, processes, etc. Should we be content with the fairly vacuous definition of ER provided by Searle or can we accept the more robust version put forward by Devitt? My view is that there's no harm in opting for the latter (though Searle thinks it is a "very deep mistake" to do so – see Searle 1995, 155). It captures the sound intuition that the realist is committed to there being a particular independent structure out there.[3] Often we get the structure wrong, but that doesn't imply that we get *no* part of it right. The progress of science, for example, is nothing but a process of getting at ever more precise representations of this independent structure.[4]

My more important misgiving concerning the pallid way Searle delimits ER is the following. Realism is not only opposed to idealism of the days long past. Nowadays the external realist has to deal with *constructivists* – quite a lot of them, actually.[5] The problem I see in Searle's

---

[3]   Hacking (1999, 83) calls realism "inherent-structurism" in order to capture precisely this idea.

[4]   Consider how Searle himself proceeds in his discussion of realism. In (1992) and (1995) he mentions the following worldly specimens: mountains, dogs, cats, stones, trees, water (snow, ice), planets, horses, fleas, hydrogen atoms, electrons and light. Doesn't his realism concern *them*? His ER and a more robust ontological commitment differ only in the level of generality, not in principle. Searle's "a way the world is" is nothing but a generalized version of "mountains, dogs, cats, stones, trees, …". I do not see why mere opting for a more specific ontology strips me of my realism.

[5]   I address a surprisingly widespread tendency towards ontological constructivism in recent American philosophy in my forthcoming book Realismus a relativismus [Realism and Relativism].

ER: there seems to be no way to distinguish it from various versions of ontological constructivism put forward in the literature; and constructivism *negates* realism. The constructivists claim that the world has no intrinsic structure. There is no way it is apart from our descriptions. It is we who divide it up into objects, properties, etc., by means of our representations. I was, to put it mildly, surprised to read the following words in Searle (1995, 160):

> The world divides up the way we divide it, and if we are ever inclined to think that our present way of dividing it is the right one, or is somehow inevitable, we can always imagine alternative systems of classification.[6]

This sounds as if the world apart from our representation were just a featureless, shapeless lump that did not contain any specific objects of properties apart from our conceptual interventions. This would be the "world well lost" of Goodman the irrealist (see Goodman 1978, 4). Such a featureless world is, I submit, the price of defining realism as being about some totally unspecific *something*.

The world is structured apart from all our representations. It contains fleas and giraffes, water and aluminum. These objects and substances are parts of it regardless whether we care to devise labels for referring to them or not. On the other hand, take *klurgs* (see Searle 1995, 160f.). A klurg is a circle randomly drawn across a portion of a book and a portion of a table the book rests on. My claim is that klurgs are no part of nature, since they do not exist apart from humans. They do not form a natural kind. In a humanless world, klurgs would not come to exist; fleas, trees and supernovae, though, would still be parts of it. Let me put it like this. Either the world contains dogs, grass and supernovae without humans and their representations, or it doesn't. If it does, we cannot divide it up in any way we please – provided our aim is to describe it correctly. If it doesn't, we may cut it up in any way we like, but then it doesn't really make sense to speak as if there was "a way the world is" apart from our representation; no *brute* facts, as opposed to socially constructed ones (e. g., that my bank account is not entirely empty), are in the offing.

---

[6] Searle (1992, 417) speaks, in a similar vein, about "a language-independent reality [that] can be carved out or divided up in different ways, by different vocabularies".

A realist in my sense is committed to the claim that the world does contain inherent boundaries, that it is not an amorphous lump. If we are lucky enough, we hit upon some of these. But we do not *create* them. Unless realism is to collapse into constructivism, we must be very clear about this. In fact, Searle seems to admit this much. According to his more cautious description of the way we employ language in cognition, we set up criteria for using our terms and *then let the world itself decide that they are empty or not* (see Searle 1995, 166). Note what this implies. The fact that the term "giraffe" does have a reference, i. e., isn't empty, *is just the fact that the world contains giraffes representation-independently.* Giraffes are denizens of the independent world, parts of its inherent structure. They did not come into being only upon the emergence of the human term "giraffe". We make descriptions, not worlds (see Searle 1995, 166).

3   Conceptual Relativism and a Privileged Conceptual Scheme

Above I, following Searle, defined conceptual relativism as a thesis to the effect that there may be indefinitely many ways of describing what amounts to the same facts. That is OK, provided we understand it in the right way. "We can measure weight in pounds or kilograms and it makes no sense to claim that only one of them correctly captures the properties of physical objects", I said earlier. In all the examples of this sort, the different descriptions are *strictly cognitively equivalent*. If one of them is true, the other is true as well, and there is a straightforward way of translating one into the other (and vice versa). Cases of conceptual relativity thus defined contrast sharply with the possibility that "the same statement (not the same sentence but the same statement) could be true of the world in one conceptual system but false of the world in another conceptual system" (Searle 1995, 167). The same portion of the world cannot correctly be described as both *p* and not *p*. ER is compatible with conceptual relativism, but it excludes the latter possibility of alethic relativism, since within the realistic framework, all true descriptions can consistently be affirmed together.

Searle further argues that conceptual relativism undermines the idea of One True Theory of the World, or, in his words, Privileged Conceptual Scheme (henceforward PCS). I confess I am at a loss to see in which way conceptual relativity interferes with PCS idea. What's wrong with

PCS? I grant that one state of affairs can be described in many different ways. But given these different ways are just *notational variants*, as we have just seen, the idea of CPS can accommodate this fact. A sensible formulation of CPS must allow for differences in alternative descriptions of the same things. But since these differences are always only superficial, we can treat the alternative descriptions simply as different ways of expressing a single PCS.[7] It seems to me that unless no stronger conceptual relativity is forthcoming, we have no reason to jettison PCS. And I don't know what this stronger relativity could amount to, were it not collapse into the alethic relativism Searle correctly rejects. If Professor Searle finds stronger relativism plausible, he should be a bit more specific about its details.[8]

So far we have seen nothing at all to cause the champions of PCS idea the least bit of unease. But realism, Searle points out, is not a thesis about *descriptions* or conceptual schemes, but a thesis about independent reality. I agree. Notice, though, that if my reasoning above is sound, the realists may commit themselves not just to a completely general way the world is, but to a much more specific ontology – and this ontology calls for a single vocabulary precisely on the lines of the

---

[7] Searle seems to grant this much in his talk about a conjunction of an indefinite number of different formulations of the same facts (see Searle 1992, 420). In a more economical vein, we could always take just one of the cognitively equivalent descriptions as a representative of the whole class. This leaves us with a finite, thought still a very large set of descriptions conforming to the PCS idea.

[8] At one place, Searle defines conceptual relativism in a way that sharply contrasts with the "official" definition: "different *and even incommensurable* vocabularies can be constructed for describing *different aspects of reality* for our various purposes" (Searle 1995, 155; emphasis added). On the official version, the schemes capture the same facts in cognitively equivalent ways. Here, the schemes address different facts by ways which are anything but cognitively equivalent. (I am not quite sure what Searle means by "incommensurable" schemes but at least it is clear that they aren't cognitively equivalent.) So, in plain words, we have one way of describing fact *a*, another, cognitively non-equivalent way for describing fact *b*. But that is only to be expected! We describe human brains in a manner appreciably different from the way we describe the social behavior of ants. All these different, cognitively non-equivalent descriptions of different portions of reality are, if true, simply different bits of the PCS. In fact, the second version of conceptual relativism seems to be no genuine species of relativism (and clearly it doesn't refute PCS).

PCS. There is thus a clear sense to the idea that the world is best de-
scribed using a single conceptual scheme, though this scheme may be
couched in a variety of slightly different terms. Such a scheme, and the
particular descriptions made by its means, is privileged in the sense
that they correctly capture the independent structure of external real-
ity. Alternative schemes are either easily convertible into it, or not cor-
rect (and thus unprivileged).

Searle finds it preposterous to suggest "that reality itself must deter-
mine how it should be described" (Searle 1995, 155). On my view, this is
precisely what reality is doing. Sure, it does not literally force us to use
the very words we are using in describing it. We can describe the world
correctly in English or in Chinese. We could adopt a global convention
to call atoms "flumps" – nature certainly does not prevent *this*. And, as
noted, we can devise novel, cognitively equivalent ways for describing
features of the world. But this is where our creativity and convention
ends, at least when we subscribe to realism. "Conceptual schemes are
human constructions and, to this extent, arbitrary" (Searle 1995, 151).
That is certainly true. And it is also true that "any true description is al-
ways more or less arbitrarily selected for describing the world" (Searle
1995, 161). But, and this cannot be stressed too much, this arbitrariness
is fairly limited.[9] I think we should resist the picture according to which
we craft our concepts on purely arbitrary grounds. Our languages do
not shape the external world. The direction of influence is very much
the opposite.[10] From the evolutionary point of view, our linguistic and
other representational capacities are molded by the inherent structure
of our environment. We just don't carve reality in any way we please.

To summarize, I have been arguing that it is advisable for the realist
to prefer ontologically more committed version of realism to Searle's
general and underspecified thesis, mainly in order to steer clear of on-
tological constructivism and other contemporary intellectual tenden-
cies that Searle (1995, 197) duly finds suspect. My second point is that

---

[9]   It is more limited in concepts for natural kinds than in concepts used for
     tracking the physical properties of our surroundings. Degrees of tempera-
     ture, for example, do not directly correspond to anything in our surround-
     ings, and this seems to be the reason why the different measuring systems
     emerged. No such conventional differences are to be found in concepts for
     animals, plants, etc. Here what varies are just their names in different lan-
     guages.

[10] In this I am in perfect agreement with Donald Davidson (1997, 16f.).

we should be careful not to read too much into Searle's denial of the Privileged Conceptual Scheme. What Searle denies is a certain literalist reading of the PCS idea which is indeed indefensible. We can, though, embrace a more sensible version of PCS that allows for superficial differences in true descriptions of the same facts. I doubt that when even this sensible version of PCS is denied, we are left with anything recognizable as "external realism".

Institute of Philosophy
The Academy of Sciences of the Czech Republic, v. v. i.
Jilská 1
110 00 Praha 1
Czech Republic
marvan@flu.cas.cz

References

DAVIDSON, D. (1997): Seeing Through Language. In: Preston, J. (ed.): *Thought and Language*. Cambridge: Cambridge University Press, 15-27.

DEVITT, M. (1997): *Realism and Truth*. 2nd ed., Princeton: Princeton University Press.

GOODMAN, N. (1978): *Ways of Worldmaking*. Indianapolis: Hackett.

HACKING, I. (1999): *The Social Construction of What?* Cambridge, (Mass.) – London: Harvard University Press.

SEARLE, J. R. (1992): Is There a Problem about Realism? *Filosofický* časopis 40, No. 3, 413-433.

SEARLE, J. R. (1995): *The Construction of Social Reality*. Harmondsworth: Penguin Press.

Searle on Emergence

Vladimír Havlík

The Academy of Sciences of the Czech Republic, *Prague*

Abstract: Searle's conception of ontological emergence is a basis for his explanation of mind and consciousness in the physical world. In this article, I try to show that a closer examination uncovers some possible ambiguities in Searle's conception of emergence. First, I try to show that Searle's distinction between emergent1 and emergent2 leads to a distinction between a strong and a weak interpretation of a causal consequence of interactions among constitutive entities and that from this point of view the existence of emergent2 is improbable only in the strong sense. Second, I attempt to clarify Searle's distinction between *explanation* and *deduction* of consciousness in his claim for the non-deducibility of consciousness. At the end I try to show in what sense is Searle's concept of emergence loaded with a form of mechanicism, one which is being abandoned in more recent ontological conceptions.

Keywords: emergence, mind, consciousness, emergent property, system property.

In recent discussions of emergence, Searle's concept of emergent properties is classified as a mainstream theory of ontological emergence. Searle uses ontological emergence as the basis for his concept of mind and consciousness, which are, for him, entirely biological properties, fundamentally dependent on their biological bearer – the brain. Mind and consciousness belong only among the properties of the whole complex system, not of the constituents of that system. In order to explain his notion of the specificity of mind, Searle often uses an analogy with physical properties (Searle 1984; 1992; 1997; 2004). He refers to physical properties such as liquidity, solidity or transparency, which can be attributed to matter only at a certain macro-level of reality while for the micro-level constituents it is impossible to embody them. The molecule of water is neither "wet" nor "liquid" and it also

lacks many other physical characteristic of water understood as a macroscopic liquid, e.g. capillarity. Analogically, consciousness is a causal emergent property of the complex brain system at the macro-level and there is no such property at the micro-level of separate neurons. The role of physical analogies is important for our understanding of emergent properties as a common thing in nature. It is usual that complex systems embody novel properties or behaviors which their constituents do not embody at elementary levels. From this point of view, the existence of mind and consciousness is nothing special because there are causal powers which lead to these emergent properties. Their distinctiveness lies in the fact that they are pure biological properties and cannot be created out of their biological bearer.

Searle's physical analogy with liquidity has often been criticized. Liquidity cannot be in the same relation to water as consciousness is to brain, because liquidity can be deduced from the properties of elementary particles whereas consciousness cannot be deduced from the properties of neurons.[1] This objection seems to me somewhat questionable. It is true that physicists could explain why and under what exact conditions there occurs the phase transition of numerous molecules of water to liquidity whereas the exact conditions for the transition of a system of neurons to the state consciousness are not known. Yet I do not think that this could be the principal point of Searle's analogy. His argument lies in emphasizing the interconnection between micro-level and macro-level, such as the emergence of liquidity of water, and this is a common physical phenomenon, manifested at many different levels of reality.

> Consciousness is a higher-level or emergent property of the brain in the utterly harmless sense of "higher-level" or "emergent" in which solidity is a higher-level emergent property of H_2O molecules when they are in a lattice structure (ice), and liquidity is similarly a higher-level emergent property of H_2O molecules when they are, roughly speaking, rolling around on each other (water). (Searle [1992] 2002, 14).

_____

[1] E.g. "What Searle fails to see is that liquidity can be predicted from the properties of elementary particles, whereas consciousness cannot be predicted from the properties of neurons." (http://www.scaruffi.com/mind/searle.html)

Drawing on this assumption, it is not surprising that a similar "mechanism" is present in the case of neurons and consciousness. Searle is well aware that we don't know the exact conditions of emergence of consciousness; he presents the mechanism of emergence as an attempt to bridge the gap in our knowledge. He expects that a similar mechanism, common as it is in many other cases, will prove mind and consciousness to be biologically and physically natural things and the gap in our knowledge will be filled in the future by neuroscience and natural sciences in general.

These physical analogies have yet another important aspect. An explanation of all the macro-level properties of water as a liquid is in most cases possible only *ex post facto*, i.e. once we know them to be manifest at the macro-level, we can investigate under what conditions they occur. However, to ante-define or predict these system properties, i.e. to deduce them exclusively from knowledge concerning properties of micro-structure (e.g. from knowledge concerning the properties or features of the water molecule), is a completely different task. Moreover we now have what I take to be compelling evidence that it is impossible. From this point of view, the emergence of liquidity of water as well as the emergence of consciousness in the neuronal structure of the brain is similarly emergent.

Searle investigates the problem of emergence in more detail in *The Rediscovery of the Mind* (see Searle [1992] 2002, chap. 5). He considers emergent properties as a type of system properties (features) that are not, or not necessarily, properties or features of elements creating the system. E.g., the shape and the weight of a stone are properties which the molecules creating the stone do not have. In fact, Searle distinguishes two types of system properties – *deducible* and *non-deducible*[2] – in a manner very similar to J. S. Mill.[3] Some system properties can be deduced, figured out or calculated from their composition and ordering, and sometimes from their relation to the environment. However, there are system properties which cannot be deduced or calculated from lower-level orderings of elements and environmental relations. This type of properties then must be explained in terms of causal interactions between micro-level elements. Searle calls them "causally emergent

---

[2] Searle does not use these terms to label this distinction.

[3] For Mill's distinction between homopathics and heteropathics laws cf. Mill ([1843] 2005, chap. VI, 242).

system features". To this type of properties belong the already mentioned macroscopic physical properties, such as solidity, liquidity and transparency (Searle [1992] 2002, 111).

Although Searle's main task is not to analyze emergence by itself but rather to use it only as a means for an explanation of the existence of consciousness in the physical world (which is an overall acceptable procedure), a closer examination uncovers some ambiguities. In the following, I will try to demonstrate them and clarify some of them, while leaving some other open. At the end an attempt will be made to show in what sense is Searle's concept of emergence loaded with a form of mechanicism.

I will start with Searle's famous claim about consciousness, what I call the *non-deducibility of consciousness*. He says:

> The existence of consciousness can be explained by the causal interactions between elements of the brain at the micro-level, but consciousness cannot itself be deduced or calculated from the sheer physical structure of the neurons without some additional account of the causal relations between them. (Searle [1992] 2002, 112)

While emphasizing the importance of the additional causal relations among entities constituting the system, he simultaneously presupposes that these additional causal relations must be explained by causal interactions at the micro-level. The requirement of explicability of these causal interactions at the micro-level leads Searle to a distinction between two types of causal emergence, which he calls emergent1 and emergent2. Searle defines emergent2 in the following way: "A feature F is emergent2 iff F is emergent1 and F has causal powers which cannot be explained by causal interactions of *a, b, c, …*" (i.e. by causal interactions of system elements – V. H.; cf. Searle [1992] 2002, 112). By this definition, the distinction between the two types of emergence is dependent on the *ability to explain* a system property by causal micro-level interactions. However, the emphasis on *the explanatory capacity* could be misleading. Searle evidently does not want the distinction between emergent1 and emergent2 to be based in the epistemological availability of an explanation. Rather, the possibility or impossibility of an explanation ought to derive from the objective state of things, and in this particular case from causal relations. Thus it is necessary to understand the distinction between emergent1 and emergent2 in the following sense: "Feature F is emergent2 iff F has causal powers *which are*

not causal consequences of causal interactions (occurring between system elements) *a, b, c, …*"

Although Searle introduces this distinction himself, he considers most emergent phenomena – including consciousness – to be of the emergent1 type. As for emergent2, he even voices some doubts whether it could exist, as he thinks that the existence of such properties "would seem to violate even the weakest principle of transitivity of causation" (Searle [1992] 2002, 112). In other words, whereas simple transitive sequence is valid for causal relations ("if event *c* is the cause of *d* and *d* is the cause of *e*, then *c* is the cause of *e*") such a principle would be invalid in the case of emergent2. In the case of emergent1, the principle of transitivity obtains, which in Searle's opinion implies that consciousness (as an emergent1 feature of the brain) is a causal consequence of neuronal interactions. Also, it follows from the definition for properties of the emergent2 type that some system properties (causal powers) are not causal consequences of interactions among constitutive elements, ergo could not be explained by these causal interactions. Unfortunately, this claim could give rise to the following misunderstanding.

First, if we adopt Searle's distinction between emergent1 and emergent2 in the ontological sense, then the explanation of system properties becomes dependent on objective causal relations. However, what is now the meaning of the claim that *a system property is not the causal consequence of interactions among constitutive entities*? What is now meant by the claim that *the system has a property which could possess causal power while this property is not a causal consequence of the microstructure*? In my opinion it is possible to distinguish between a strong and a weak interpretation of this causal consequence. In the strong version, this claim means that the examined property is not dependent on the state of its microstructure, and given that the system is in some state *S*, it *may have* but *need not have* this property. The presence of a system property with regard to the system's microstructure is arbitrary. In the weak sense, the claim means that the property is dependent on microstructure (i.e. given that the system is in some state *S*, it either *does have* or *does not have* the property) yet it is not deducible or calculable from the microstructure. In this case, the presence of the system property is not arbitrary and, in a sense, it is necessary. Thus the system property *P* is not a direct causal consequence of the only system's microstructure, but it is causally dependent on it.

Unfortunately, Searle neglects to take account of the possibility that there is a strong and a weak interpretation of causal consequence. He tends to think that any property which is not the causal consequence of microstructure is always emergent2 and its existence is improbable. However, once we admit the distinction between the strong and weak interpretations of causal consequence, it appears that the existence of emergent2 is improbable only in the strong sense, while in the weak sense it is not only highly probable but in many cases even emergently evident.

However, Searle uses a similar distinction between *explanation* and *deduction* when he claims (in the passage quoted above): "The existence of consciousness *can be explained* by the causal interactions between elements of the brain at the micro-level, but consciousness *cannot itself be deduced or calculated …*" (Searle [1992] 2002, 112, my emphasis). While the distinction between *explanation* and *deduction* may seem somewhat bizarre due to the Hempel's symmetry between *explanation* and *prediction*, the fact is that with emergent entities, there is, in practice, a temporal asymmetry between the two terms. We can *explain* something once it occurs (*ex post*), whereas we must be able *deduce* something even before it occurs (*ex ante*). What Searle wants to stress is the impossibility to know in advance which mutual causal relation between entities occurs and which properties (features) will emerge. However, emergent properties (features) which already occur and can be investigated, could be explained by the interaction of entities at the micro-level.

Second, the distinction between emergent1 and emergent2 depends on whether given property is a *causal consequence of interaction among elements or not*. In his claim for the non-deducibility of consciousness, Searle says that consciousness is not deducible "*from the sheer physical structure of the neurons without some additional account of the causal relations between them*." He thinks that the dynamics of mutual causal relations of the constituting system is crucial for the existence of system emergent properties (in the sense of emergent1). Given the fact that consciousness, as a case of emergent1, is the causal consequence of neurons, is there any possibility to deduce it from their sheer physical structure and an *additional account of the causal relations between them*? It remains an open question whether Searle would accept that consciousness could be deduced from mutual causal interactions among neurons in the brain. While he claims that this is not possible without these causal relations, he also never says if it is possible at all, and in fact

he knows that it is not. We cannot deduce or calculate consciousness either from the sheer structure of neurons or from their mutual causal relations. I think that in speaking of an additional account of causal relations, Searle means not only mutual interactions between individual neurons but also those causal relations which occur only at the macro-level and need not even be recognizable at the micro-level. Although Searle never says this explicitly, I think he would never admit that the non-deducibility of consciousness impacts attempts to deduce consciousness from sheer structure only, and not from causal interaction as well. If this conclusion is correct, then the term "additional causal interactions" is a sort of mysterious incantation whose content is only to be speculated about.

However, there is a possibility that someone wants to connect non-deducibility of consciousness with an argument for irreducibility of consciousness (the fundamental part of Searle's conception). He can then claim: "If there is something irreducible to its microstructure, then it cannot be deducible from it too." It does seem to be a very seductive possibility. Yet unfortunately, it is, once again, not evident or convincing in the case of emergent phenomena. There are features that can have causal consequences at macro-level, could be calculable by simulation but at the same time cannot be fully reducible to its microstructure.[4] While I do not have sufficient space here for a detailed argument, I do think that irreducibility cannot support non-deducibility.

I will now try to show that Searle's conception of emergence includes a form of mechanism, one which is being abandoned in the more recent ontological conceptions. Searle presupposes that entities of microstructure, such as molecules or neurons, are identical both as isolated entities and as entities involved in the establishing of the system as a whole. All that can change is the mutual causal relation, while entities by themselves stay rigid. Causal interactions between entities are the only cause and the only source of system properties or system features. Although Searle never analyzed these relations in detail, it is plausible that he would accept the following schema, founded on the transitivity of causality: 1) entities never lose their identity (such as shape, structure, features) and they manifest it even when involved in the establishing of a system; 2) entities and their properties or features are the cause of the causal relations which occur among them in the system; 3) thus,

_____

[4]   E.g., the so-called weak emergence (see Bedau 1997).

these causal relations are the cause of the emergence of system properties. System properties are in principle explicable, yet at the same time system features and the system structures are not (directly) deducible or calculable from a mere knowledge of the constitutive entities. Based on the principle of transitivity of causal relations, we must assume that the sources of consciousness are causal features at the micro-level (i.e. electrochemical states of neurons in the brain). Emergence is a common mechanism in our world, and emerging novel entities such as substances, properties, relations and behaviors are thus situated at a higher system level from the original level of its constituents. This universal scheme of emergence is realized at various hierarchical levels of reality and it is responsible for the emergence of complex systems, from chemical compounds via life all the way to consciousness. Although Searle would probably agree with this scheme, he assumed that his concept of emergent1 is sufficient for an explanation of all emergent phenomena including consciousness. However, it is possible to demonstrate that there exist emergent physical phenomena for whose explanation the emergent1 concept is insufficient due its form of mechanism. Many phenomena in physics demonstrate that entities are a different thing when they stand alone and when they take part in the creation of a system. These evidences are frequently brought up not only by physicists working on solid state physics and condensed matter physics (e.g. Anderson 1972, Leggett 1987, Laughlin 1998, Healey 2010). Recently, research in quantum theory has also provided similar evidence (e.g. Healey 1991, Silberstein – McGeever 1999, Kronz – Tiehen 2002, Hüttemann 2004).

We could maintain consistency with Searle's assumption regarding the identical mechanism of emergence of novel entities in nature, including consciousness, even in the light of the new concepts of emergence, such as the concept of "fusion" which claims that constitutional entities change radically when taking part in the creation of a system. What leads to the emergence of novel system features or properties is a fusion of elements together with their mutual causal interconnections and their interactions with the environment. We could claim that the same mechanism is responsible for, or at least present at, the emergence of consciousness, and that Searle's vision of emergence as a general mechanism is thereby not weakened but rather enhanced.

Institute of Philosophy
The Academy of Sciences of the Czech Republic, v. v. i.
Jilská 1
110 00 Praha 1
Czech Republic
havlik@flu.cas.cz

References

ANDERSON, P. W. (1972): More is Different. *Science* 177, 393-396.

BEDAU, M. A. (1997): Weak Emergence. *Philosophical Perspectives* 11, 375–399.

HEALEY, R. A. (1991): Holism and Nonseparability. *The Journal of Philosophy* 88, No. 8, 393-421.

HEALEY, R. A. (2010): Reduction and Emergence in Bose-Einstein Condensates. *Found. Phys.*, DOI: 10.1007/s10701-010-9481-8.

HÜTTEMANN, A. (2004): *What's Wrong with Microphysicalism*. London: Routledge.

KRONZ, M. F. – TIEHEN, T. J. (2002): Emergence and Quantum Mechanics. *Philosophy of Science* 69, 324–347.

LAUGHLIN, R. B. (1998): Fractional Quantization. *Nobel Lecture.*

LEGGETT, A. J. (1987): *The Problems of Physics*. New York: Oxford.

MILL, J. S. ([1843] 2005): *System of Logic, Rationative and Inductive. Being a Connected View of The Principles of Evidence and the Methods of Scientific Investigation*. New York: Harper and Brothers, Publishers.

SCARUFFI, P. (2006): *The Nature of Consciousness. The Structure of Life and the Meaning of Matter*. Omniware.

SEARLE, J. R. (1984): *Minds, Brains and Science*. Harvard University Press.

SEARLE, J. R. ([1992] 2002): *The Rediscovery of the Mind*. Cambridge: the MIT Press.

SEARLE, J. R. (1997): *The Mystery of Consciousness*. New York: The New York Review of Books.

SEARLE, J. R. ([2004] 2007): *Freedom and Neurobiology. Reflections on Free Will, Language, and Political Power*. New York: Columbia University Press.

SILBERSTEIN, M. – McGEEVER, J. (1999): The Search for Ontological Emergence. *The Philosophical Quarterly* 49, No. 195, 182-200.

II. Philosophy of mind

Sentience, Awareness, Consciousness

Martin Pokorný

The Academy of Sciences of the Czech Republic, Prague

Abstract: The paper starts from a Searlean dilemma – we are bound to view consciousness as ultimately explicable by scientific means, yet science appears to give us no means for explaining the specificity of consciousness – and presents what I see as a plausible though speculative story for avoiding the brunt of the dilemma. The basic idea is (a) that consciousness, or anticipations of it, should be seen as pervasive throughout the biosphere; (b) that the biosphere, following Gerald Edelman, can be seen as the sphere of meta-systems irreducible to purely physical particles and forces; and (c) that it is plausible to view "full waking consciousness" as occurring at a very high level of meta-systematicity; with the conclusion (d) that full waking consciousness is both an expectable outcome of the biogenic forces and, in virtue of how it combines them, a very singular case.

Keywords: sentience; awareness; consciousness; recognition systems; Gerald Edelman.

1

John Searle's basic stand regarding consciousness[1] can be characterized fairly briefly: we must take science seriously, and we also must take consciousness seriously. Taking science seriously implies the conviction that consciousness must, at least in principle, ultimately lie within the ambit of scientific research. Taking consciousness seriously implies the conviction that at the current moment we are very far from actually possessing a scientific explanation of consciousness, or even merely a detailed outline of it.

---

[1]   Cf. especially Searle (1992) and Searle (1997).

This essay is intended as a very small contribution to the effort to resolve the dilemma. I will not comment on Searle's own arguments nor quote from his texts. Instead, I will formulate a position which I take to be *bona fide* in the Searlean spirit. I take the considerable liberty of exploiting the format of this volume to the full and leaving it to John Searle himself to state to what degree these suggestions might actually be acceptable to him.

2

A few preliminary observations: First, if the brain is taken to be the most advanced product of nature's biological engineering, and if consciousness is taken as a property of the brain, then it needs to be proclaimed very definitely that present-day science is nowhere close to explaining or replicating biological systems of comparable complexity. The clash between the advocates of scientism or eliminativism and those who oppose these reductive strategies does not really concern present-day research and technology. The main thrust of the polemics is about where our research and our technology could or could not take us, along a reasonably continuous route, a *very long* time hence.

Thus, second, it might seem peculiar that disagreement about something so distant should awaken so much interest and provoke such strong emotions. And here it will be helpful to point out a fact which all the concerned surely see yet are wary to declare openly: the polemics is, at the current stage, primarily – though not exclusively – a *cultural* one, which is to say, it regards a difference of *mindsets*. The reason why this tends to remain unsaid is, obviously, the need to keep the realm of serious, primarily academic discussion separate from the "culture wars", i.e. the various excesses targeting science from a standpoint of fundamental rejection. Still, it needs to be admitted that the differences between the two culture clashes, while immense, are not total: there *are* certain continuities between the two discourses, and it is better that these overlaps, small as they might seem relative to the contrasts, be articulated exactly, in order to see their limits.

It is rather evident that the proponents of scientism and eliminativism view consciousness as the last vestige of old superstitions: consciousness is human, subjective and, globally, very scarce, and it is imperative (according to this camp) that it be swallowed by a broad and objective description, just as the Renaissance eliminated the uniqueness

of Europe, Copernicus eliminated the centrality of the Earth, Darwin eliminated the uniqueness of humankind, Einstein broke through the fundamental barriers of our sense-perception and Bohr finalized the destruction of the fundamental barriers of our conceptual thinking. To insist that the simple fact of consciousness has somehow escaped this triumphal march must seem, from the perspective sketched here, an obscurantist and regressive position.

I believe a fairly simple counter-measure needs to be taken. Copernicus, Darwin, Einstein and Bohr were builders of theories, and theories are impossible in a world without consciousness. The triumphal march of science *is, ipso facto*, a triumphal march of consciousness. The contrast is not between the microscopic and macroscopic infinities of the Universe and the scarce, discontinuous, weird and doubtful phenomenon of consciousness. Rather, the contrast is between the inhuman immensities of the Universe and the human theory-building capacity to comprise them – and this presumed contrast is actually quite close to a symmetry, even though it is not a simple symmetry. It may well be difficult to describe it in epistemological detail, but it is evident and indubitable, and it should be the keystone upon which the proponents of consciousness rest their case.

3

At another level, the advocates of consciousness ought to acknowledge a continuity between high-level consciousness and low-level sentience and proclaim the ubiquity of conscious and/or sentient phenomena throughout the biosphere. If, perhaps, we lack logically apodictic evidence for attributing (some form of) consciousness and/or sentience to horses, cows, lizards, birds, fish, plants and bacteria – and I am not sure that we lack such apodictic evidence, since I am not sure what it would be to possess it –, then it is a cognitive gap we should mark and register, but it gives us no reason for actually denying them either sentience or consciousness. While biotic phenomena may be scarce in the Universe they are not at all scarce on the Earth, and within the ambit of the sciences of life, consciousness and sentience – phenomena different but close – are simply all over the place.

I suspect a strong influence here is the assumption that the progress of science stands in a direct proportion to the gradual demise of anthropomorphism. However, in European intellectual history all critique of

anthropomorphism had a certain primary field, which was theology; and while it is true that, generally speaking, the progress of science went side-by-side with the theological tendency away from full-throttled anthropomorphism and nearer the concept of a purely rational, mathematical God, this implies next to nothing about what we should expect once we simply leave religion aside. As a matter of fact, after Darwin we actually have a kind of obligation to be anthropomorphic with respect to the biosphere, since the concept of evolution obligates us to assume that whichever capacities we find in ourselves are to be expected – in some recognizably comparable form – along an indefinite network of evolutionary lineages. These initial expectations can certainly be disproved and frustrated by detailed research, but even so it remains legitimate, or even imperative, to adopt them as a preliminary starting point. There is no good basis for the notion that methodological skepticism with regard to consciousness and sentience in species other than humans should be the default scientific attitude.

4

Thus I take it that consciousness is a robust fact of nature: it is – either by itself, or in the anticipatory form of sentience – massively present in the biosphere, and it constitutes the cognitive form of all those vast stretches of the Universe that lack it materially. Even though present-day science might not possess a full-fledged account of it, it is plausible to expect that with the current state of our knowledge we should at least be able to articulate the principles which could reasonably lead to an explanation of the phenomenon.

In other words, consciousness by itself is not an urgent problem of current scientific research. What is urgent is the puzzle facing the scientifically minded whether and how consciousness can be, *in broad strokes*, incorporated in the picture of nature which we have attained.

I want to outline a defense of the claim that the principle of such broad incorporation – or a good candidate for it – is contained in Gerald Edelman's concept of *recognition*.[2] Edelman defines it as follows:

---

[2] For this concept in particular, I draw on Edelman (1992). The necessary background is, of course, constituted by Edelman (1987), Edelman (1988), and especially Edelman (1989).

> By "recognition", I mean the continual adaptive matching or fitting of elements in one physical domain to novelty occurring in elements of another, more or less independent physical domain, a matching that occurs without prior instruction. (Edelman 1992, 74)

A prominent example of a recognition system is the immune system: cells that carry antibodies multiply in proportion to the use found by the particular antibody molecules they carry. Edelman explains:

> It exists in one physical domain (an individual's body) and responds to novelty arising independently in another domain (a foreign molecule among the millions upon millions of possible chemically different molecules) by a specific binding event and an adaptive cellular response. It does this *without* requiring that information about the shape that needs to be recognized be transferred to the recognizing system *at the time when it makes the recognizer molecules or antibodies*. Instead, the recognizing system *first* generates a diverse population of antibody molecules and then selects *ex post facto* those that fit or match. It does this continually and, for the most part, adaptively. (Edelman 1992, 78)

Besides immunity, the other major types of recognition include heredity, neural reflexivity and the brain capacity of categorizing and recategorizing (see Edelman 1992, 205).

Edelman stresses that in all this there is "no final cause, no teleology, no purpose guiding the overall process" (Edelman 1992, 74) – which is correct to a very large degree. However, a certain dose of qualification is needed here. While Edelman properly avoids positing any *particular* goals, he does elevate recognition (in the sense defined above) to a grand principle of nature, parallel with the principle of (physical) symmetry. By this he means that the establishing of recognition metasystems (as, by definition, they have to include at least two relatively independent sub-systems) is a process irreducible in its entirety to physical particles and forces. For instance, it is a major goal of Edelman's *Topobiology* (Edelman 1988) to point out that even though the development of the embryo certainly *obeys* the laws of physics, it is not reducible to them: embryogeny requires topological attraction and differentiation which is specific to the interactions within the cell population that eventually produces the embryo, even though the cells are not "instructed" toward a particular outcome, i.e. they do not "contain an information".

All this equals saying that there is another force of nature besides the forces we would normally call "physical", a force which we can observe in biotic processes. As I see it, the most cautious way of formulating this is to say that *there is a tendency in nature which, from a certain threshold, becomes observable as a biogenic tendency*, with "biogenic tendency" standing for a tendency to produce meta-systems of recognition in Edelman's sense.

What do we gain by postulating such a principle? We gain the possibility to state that there really are organisms,[3] and that there really is consciousness, while retaining the lessons of Darwinism and while attaining a very large degree of continuity between the purely physical and the biotic.

Does, then, modern science give us a clue to understanding consciousness? It does and it doesn't. The *phenomena* described by biological chemistry, embryology and neurobiology contain the principle needed for achieving a rough integration of consciousness within the present-day, science-informed worldview. However, it is one of Edelman's points that the pertinent sciences attempt to limit themselves to a purely physical description, thus leaving an explanatory gap. Edelman seeks to fill the gap. The result is a more complete description, one which works with natural recognition as a factor. The so-called principle of recognition, or the biogenic tendency, is, then, not formulated in the abstract, and perhaps that is just as it ought to be. It is sufficient that it can be *read off* the analogy between immunity, heredity, neural categorization and other, similar phenomena.

5

However, immunity, heredity and neural categorization are non-conscious. How are we to get from recognition systems to consciousness?

One part of the answer I want to propose is: in the workings of the brain, the biogenic tendency establishes itself in multiple layers,

---

[3] To be sure, I am not proposing any concept of organism in this essay, nor is it contained in Edelman's books. But I take it that it is sufficiently plausible to assume that with due effort expended, the principle of recognition would allow us to provide a characterization of organism, or something substantively analogical to the traditional notion of organism, as a natural, inherently consistent entity.

or loops, to a degree unobservable anywhere else in nature; and this makes it plausible to state that consciousness is something very special *and at the same time* the outcome of a general natural principle.

I will list five *prima facie* plausible mechanisms of such layering. The first three can be called mechanisms of established or retained complexity. It is orientation, perception of an object, and proprioception.

Orientation means that a material stimulus appears against a system of dimensions (or a single dimension) in practical space: one and the same X can be near or far, within grasp or out of grasp, at the protected side or at the weak side; and depending on the attractive or repulsive character of the X, the organism acts and moves so as to keep it closer to one or the other extreme of the dimension. This means that orientation requires a coordination of at least two relatively independent systems, i.e. the system of material inputs and the system of orientation.

Perception of an object means that stimuli are inserted into a thing-system, with the present stimuli complemented by expected or excluded stimuli. We can leave moot the question whether these stimuli could be all of the same type, and simply state that typically, in higher organisms, they will be of various types: for instance, the same thing can be seen and smelled.

Proprioception denotes, in brief, the capacity of the organism to sense itself; or more cautiously, the capacity of the organism to have a wealth of stimuli regarding the tissues and organs that make it up.

It appears almost evident that the last capacity builds on the previous two: differentiated proprioception requires orientation, unified proprioception requires that the organism's body (or more precisely: the proprioceptive part of it) be perceived as an object – and proprioception is nothing unless it is both differentiated and unified. So it might be seen as a kind of *product* of the previous two capacities. But we have also reasons to think that the relation is reciprocal: (a) the development of proprioception opens new avenues for the development of orientation (therefore, it does not seem an accident that all the animals that manifestly share our experiential space also possess highly differentiated and subtly proprioceptive bodies); (b) one's body is the object the organism will come to know best, and in the most detailed manner, so that the organism's "subject" actually is and remains, in terms of long-term salience, the first object.

Thus, there is strong commonsensical evidence that practical space, objects, the several senses, and the body zone (as we can call it if we

want to avoid a too early mention of "self") are systems of input which coordinate and develop via mutual stimulation yet also remain independent. In Edelmanian terms, this means that there must be multiple coordinated systems of recognition: the neural re-entry loops must cross without blending.

The other two mechanisms on my list can be termed simplifying mechanisms. I will call them awareness and consciousness, diverging from both Edelman's and Searle's use.

Let me characterize awareness as the capacity to pass from one regime of sensitivity to another regime of sensitivity. By regime I mean that the organism, in its channels of sensitivity, shifts emphasis *not* from one sensation or perception to another but rather from one range of sensations or perceptions to another, or differently tuned, range.

And let me characterize consciousness as the memory of awareness. The crucial point is that awareness and consciousness are capacities that concern capacities: awareness is the capacity to "switch on" or "turn up" this or that sense-channel; and consciousness is the capacity to remember awareness, i.e. a capacity.

Awareness, as I define it, is probably fairly close to Edelman's primary consciousness, and it would be included in Searle's use of the term "consciousness" without attribute. However, the point of my definition – if one accepts that it corresponds to something real in our human experience, as well as in the experience of higher mammals and perhaps other animals – is to bring out that it is an economizing mechanism: an organism has only a very limited range of senses; and awareness provides active access – a *unified* access – to their variety. Awareness can broaden its range or its subtlety, but in an important sense there is *just one*.

Consciousness, as the memory of awareness, then makes awareness thematic: every particular momentary modality of awareness appears against the background of awareness *tout court* – of all the other regimes that the organism might employ.

If we add up the simplifying mechanisms to the mechanisms of established or retained complexity, and understand them all as workings of neuronal re-entry, we get the idea of multiply crisscrossing yet distinct loops *topped over* with a re-entry mechanism related to entire large groups of re-entry mechanisms, a kind of directing center (i.e. awareness), *plus* yet another re-entry mechanism establishing a memorial loop for this directing center (i.e. consciousness).

To sum it all up and draw the conclusion: if we acknowledge the principle of recognition, or the biogenic tendency, as a real force of nature, then we have strong *prima facie* evidence to think that in highly developed brains, this force is combined, re-combined, and meta-combined to a degree not observable anywhere else in nature. As a result, consciousness appears *both* as a very special case *and* as a "logical", i.e. non-surprising consequence of a broadly valid principle of nature.

6

Let me answer three objections that are certain to come up: first, the argument is circular – I explain consciousness by assuming it and working towards it; second, I do not explain why consciousness "feels" the way it feels; and third, the whole argument is too speculative to be of any scientific value.

For the first objection, that the argument is circular: That would be true if I was trying to demonstrate that consciousness exists. However, the existence of consciousness I hold for evident, and don't try to prove it: consciousness is a given. What I am attempting to show is the explanatory compatibility of the phenomenon of consciousness with a certain rational view of nature. *This* I did not assume; it is a pure outcome.

For the second objection, that I do not explain why consciousness "feels" the way it feels: Consciousness is an evolutionary product of nature, and we simply do not have the option to second-guess nature and ask whether she could have proceeded otherwise: whether lions could have been made more gentle, eagles less hungry, and human minds somehow different in their feel. What we do try to explain is the evolutionary purposefulness – or precisely, the fact of their having been selected for – of the way things are. This is easy on the outward side of my argument: perceiving more, and having more control over one's capacities of perception, is – *ceteris paribus* – a clear evolutionary advantage. Where it is difficult is on the inside: how come that the brain can evolve in this way. And here I simply draw on Edelman's neural Darwinism, and on the principle of recognition it employs as a factor: if the functionality of even the simplest neural system is based on the capacity of re-entry, then it is not surprising that the further evolution of brains as neural centers should involve not only a more extensive

use of this capacity but also a structurally more complex use of this capacity.

For the third objection, that the argument is too speculative, moving in a realm too far away from any scientific theory of consciousness: This I will answer in two parts. First, it is not speculative, it is merely sketchy; and it is okay to rest content with a sketch if a sketch is all we need; and since my only goal was to point out – with more emphasis than I have found it said in Edelman – the multi-layered and multi-looped character of any complex experience *and* of the brain corresponding to it, I can rest content with a mere outline, confident that even if it stood to be corrected it is unlikely to become measurably simpler. Second, any scientific theory of consciousness *has* to start from some commonsensical divisions and descriptions, and it will never completely overcome this starting point. It is true that, from this viewpoint, I present here a list of starting points and "no beef"; but that does not make the starting points somehow illegitimate – and given that I have no competence in neurology whatsoever, it is only so well that I do not attempt to speculate about the possibilities of empirical research.

These are quick answers, not meant to convince the skeptic but rather to clarify some *prima facie* puzzles to a reader thinking along substantially the same lines.

7

So far I have stressed merely the overall correlation between the complexity of consciousness (or of its structural assumptions) and the complexity of the brain. Now it is time to get to the specifics.

I find my proposal attractive because I think it can give a plausible overall picture for an explanation of some essential features of sentience and consciousness.

Let me list some of the ideas the model suggests:

(a)   If we accept that the motor of the evolution of sensation and perception is the interaction between orientation, the perception of an object, and proprioception, it suggests an idea why certain stimuli "rise up to full awareness" and others don't. Why are we aware of our sensations but not of the detailed workings of our immunity system, or more plausibly, our visual or tactile system? My answer: because one has become involved in the interaction mentioned above, and the other hasn't; one should resist the temptation to look for further rea-

sons. There is a certain minimum of looping complexity, quantitative and qualitative, which has to be attained in order for awareness to find purchase; and it seems to be at, or near, the crossing point of the three "lower" mechanisms listed above.

(b)　The three "lower" capacities have certain structural constants: proprioception keeps a focus on the body occupied by the organism (as I put it awkwardly to avoid any mental expression), orientation relies on certain pervasive dimensions, and the perception of an object is anchored in the (evolving and flexible yet always characteristic) form of an object. Awareness, as I have defined it, is also substantially an awareness of their permanent cooperation. Thus, it is plausible to expect that an organism's awareness will have a structural constant resulting from this cooperation: an operative center of the principal dimensions of practical space, constituted by an object delineated by the proprioceptive area. Once available to awareness, this will already deserve to be called a self – emphatically, a self available to certain animals, too.

(c)　Consciousness, as I have defined it, is a step above awareness: it is the memory of awareness. Awareness is, so to say, one in itself but turned to many; consciousness is one in itself and turned to one. Awareness is readiness to modify one's attitude to the situation, consciousness is the aggregated memory that every attitude to every situation is simply a case that can be modified. Thus, it is homogeneous, continuous, unitary.

(d)　What is a plausible story about the evolution of consciousness out of awareness? I think there are two basic options: either a great ape wakes up one morning and finds itself self-reflective; or else an awareness of self (in the sense given above) passes over into an awareness of other. I plead for B. As a matter of fact, it only says that awareness is prone to spill over from the area of proprioception to the area of perception of an object (or, possibly, vice versa – which comes to the same for our present purposes), and that it colors the dimensions of orientation. Thus, objects given in experience can start to "behave" thus or thus, and therefore to have selves. If interaction of self with selves becomes non-sporadic, consciousness arises.

(e)　If we accept the previous two points then it follows that the awareness of awareness is consciousness but the consciousness of consciousness is still consciousness. Of course this is partially just a matter of definition but my point is that it makes sense to draw the line this

way. Edelman believes that higher-order consciousness is created by language. My question would be, how can an animal possibly acquire the symbolic capacity if such a capacity as of yet completely transcends its horizons? Thus, I believe a more productive – and *more naturalistic* – formula is to set language in between consciousness and consciousness of consciousness. The speaking animal has all it needs to employ language, i.e. symbols, it only needs to find a suitable reservoir of them and to start using them; however, the encounters with (various employs of) language, once it is acquired and its use expanded, make the animal gradually conscious of what it had to have *in order* to acquire and effectively employ language. This starts a new round of interaction among the animal's capacities, one which is presided over by the imperturbable form of consciousness but which actually consists in new and even more productive combinations of orientation, objective perception, proprioception, awareness-as-readiness, and objectified selves.

8

In conclusion: I have started by characterizing the debates surrounding consciousness as, in important respects, debates *about* science rather than *within* science, and then debates about *general principles* rather than substantial particulars. In line with this, I have found it appropriate to present merely a likely story. The gist of it is that the human brain is the most complex yet somehow unitary arrangement of living tissue, and that this is what enables it to anchor the phenomenon of consciousness. I have used Gerald Edelman's neurobiological concept as a basis which has enabled me to spell out the manner of this complexity-in-unity. There may be other, very different neurobiological approaches that could fulfill the same role. The key point is that in order to do that, they have to understand themselves as sciences of life, not *exclusively* as sciences of electromagnetic resonances. Again, I have proposed the "biogenic tendency" as a force of nature whose prominence delimits the realm of life from the realm of non-life, and again I relied upon Edelman's interpretation of immunity, heredity and other processes of "recognition". It is up to specialists to decide whether Edelman's statements here might stand corrected, and in such a case, what I had to say about the "principle of recognition" would stand corrected too. But the crucial point is: corrected, not eliminated. It is on this basis that I feel

reasonably confident that the story, though merely likely, should contain an element of truth.

Institute of Philosophy
The Academy of Sciences of the Czech Republic, v. v. i.
Jilská 1
110 00 Praha 1
Czech Republic
pokorny.cz@gmail.com

References

EDELMAN, G. (1987): *Neural Darwinism*. New York: Basic Books.
EDELMAN, G. (1988): *Topobiology*. New York: Basic Books.
EDELMAN, G. (1989): *The Remembered Present*. New York: Basic Books.
EDELMAN, G. (1992): *Bright Air, Brilliant Fire*. New York: Basic Books.
SEARLE, J. (1992): *The Rediscovery of the Mind*. Cambridge, (Mass.): MIT Press.
SEARLE, J. (1997): *The Mystery of Consciousness*. New York: New York Review of Books.

Causality and Free Will

Juraj Hvorecký

The Academy of Sciences of the Czech Republic, Prague

Abstract: A comparison of Searle's contrasting position on explaining consciousness and free will is conducted with an aim to show that while consciousness on his view presents a relatively easy problem, while free will is presented as a serious challenge to naturalism. We suggest that with a certain shift of perspective on causality, free will problem might be just as tractable as the problem of consciousness.

Keywords: consciousness, free will, determinism, causality.

On Searles's account, determinism presents a serious challenge to our attempts to uncover the relationship between mind and matter. In this brief paper, we will concentrate on his outline of two possible answers to the issue, as presented in his widely read paper *Free Will as a Problem in Neurobiology,* and present some recent empirical evidence that seems to be lending some support to one of them. Along the way we will try to pinpoint few questions that Searle is taking quite seriously, while there might be good reasons to leave them aside, because they create more problems than they solve.

The problem of free will, as described by Searle, is indeed very well known. We have two opposing intuitions about the nature of actions. On the one had, "in our dealings with nature we assume that everything that happens, occurs as a result of antecedently sufficient causal conditions" (Searle 2001, 495). All events have their prior causes. The nature of the world is deterministic. Recent developments in quantum physics give us only little hope of bringing indeterminism into the physical picture of the world. And even if physical indeterminism is assumed, fitting it into the overall picture of perceived free action is far from clear. On the other hand, we understand ourselves as *causa sui,* initiators and executors of our own will and free agents. The main source of this self-understanding consists in permanent perception of several stages of

gaps in our consciousness of action, from its initiation all the way to its execution: "I do not sense the antecedent causes of my action in the form of reasons, such as beliefs and desires, as setting causally sufficient conditions for the action" (Searle 2001, 493). These gaps in perceived actions cannot be simply bridged by an assumed physical indeterminism and no other bridge is available. The clash between causally sustained chain of events in the world and apparent gaps in actions initiated by us is what constitutes the essence of the problem of free will.

Searle, instead of simply stating the opposing intuitions and fully in line with his biological naturalism, tries to accommodate experiences of freedom with the underlying naturalistic framework of the mind. He has repeatedly argued that mind and its conscious components are nothing but causal consequences of lower-level micro processes in the brain. To use his famous phrase, the brain *causes and realizes* higher order mental states. His solution to the mind-body problem is uncomplicated, yet elegant and we are fairly sympathetic to its general line. Given our liking of the overall Searlean model of the mind, we find the way one of the central question of the paper is posed a bit troubling. Searle's question is precisely "what would the behaviour of the neurons and the synapses have to be like if the conscious experience of free will were to be neurobiologically real?" (Searle 2001, 503). In other words, he is interested in "how might [the] gap be reflected at the neurobiological level" (Searle 2001, 504). He then goes on to suggest two competing hypotheses on neurobiological underpinning of volition. Before briefly commenting on those two hypotheses, we would like to make a general point, regarding the very question of a relation between neurobiology and freedom. We believe it is of some relevance that an analogous question on a relation of neurons and synapses to consciousness has recently produced some highly undesirable philosophical developments. While asking what the behavior of neurons and synapses have to be like in order for consciousness to appear, several authors have come to the conclusion that they see no way to detect any property in the world of science that would differentiate neuronal activity, responsible for non-conscious mental states from that which underlies conscious mental states. A pessimistic conclusion that none of neuronal properties that we are likely to discover allows for an explication of special experiential character of conscious states often leads to speculative and unsubstantiated claims that uncovering the nature of consciousness will require transcending results of biological or any other natural sciences.

To some (Chalmers, G. Strawson) this skeptic road inevitably leads to the abyss of panpsychism, with elementary psychological properties posed at the fundamental level of micro-particles. Panpsychism therefore raises from a desire to explain the difference between conscious and unconscious states and a belief that it has to be reflected in some phenomenal property at the lower level.

Searle is far from being lured by this line of argumentation. Instead, his solution to the mind-body problem has always been straightforward and decisive. He has repeatedly argued that all mental states are caused and realized by the brain. As we have seen, the insistence of various authors on the reflection of phenomenal properties at non-psychological level lead to claims of panpsychism. Searle and many others think this is completely unfounded misunderstanding of how properties are assigned at various levels. We believe it is the very question of *what lower level properties have to be like* in order to produce consciousness that lead to overall rejection of physicalism among those who insists the answer has to invoke a mirror between lower level and higher level properties.

Searlean quest to detect behavior at the neurobiological level that mirrors high-level phenomena of free volition shows some apparent similarities to the quest of panpsychism. It asks what properties at the lower level correspond to our experiences of gaps. But why assuming that gappy properties of the higher level have to correspond to gaps at the lower level? We have seen that an insistence of discovering analogous properties for conscious states opens up the door for a Metaphysical Zoo of panpsychism (to borrow a phrase from Russell) with unorthodox properties introduced into general ontology. There is a danger that Searle's query to locate gaps at the lower level might end up with an introduction of several unorthodox non-physical properties that go against his proclaimed biological naturalism.

The issue is especially striking given the Searle's insistence that consciousness in the brain is localizable and that science will tell us more about how, where and why it occurs. There is no further philosophically interesting issue to be pursued there. So it is not easy to see why Searle asserts a need for an additional neuronal foundation for volitional acts, when the evidence for them comes solely from our conscious perceptions and these in turn are caused and realized by appropriate neuronal states. Why to expect an existence of a corresponding gap among lower level mechanisms when evidence for the gap is psychologically robust, but there are no gaps to be discovered anywhere in the brain? If

Searlean answer to the problem of emergence of conscious states is that brain simply causes them, it seems equally sufficient to claim that experiences of freedom are also caused by relevant neuronal circuits and no further philosophically intriguing questions are to be found here.

Yet the worry about free will seems to persist, because experiences of freedom are constitutive of our self-perception as free agents and beings responsible for our own actions. So if our answer to Searle's inquiry into how gaps are reflected at the neurobiological level is that our brains work in such a way that we are prone to have perceptions of free action while in fact all sufficient antecedent causes of all our actions are handled by the brain in a purely deterministic manner, we have already opted for the Hypothesis 1 and its resulting epiphenomenalism.

Let us point out that recent years has seen some interesting new scientific evidence for the epiphenomenal claim that seem to bring new support for the Hypothesis 1. Most striking of all is probably (Soon et al. 2008) where subjects asked to view a line of numbers on the screen and at the moment of their free choice have to push a button with either their left or right hand. Then they are to report the number which was on the screen when their decision was made. These first-person data are then compared with fMRI scans of a particular brain area and a correspondence of the brain activity with the decision is discovered. Shockingly, a brain event up to 7 seconds prior to subjects' decisions was found to correspond with a supposed free act and, based on this correspondence, a prediction about which hand is about to be used could be made very early on. Authors claim that given a relatively slow response speed of fMRI scanners, the actual neuronal event could be present in the brain almost 10 seconds prior to the subject's decision. Let us also add that the same experiment was replicated last year by another team with a more efficient scanner (Bode et al. 2011). To our knowledge this seems to be the most striking scientific evidence for epihenomenalism so far. It should come as no surprise to any naturalist that our free actions are preceded by some neural activity or other very briefly before an onset of an action. However, seven seconds is indeed a very long time and if similar findings are to be demonstrated in other domains of our activities, consequences would be disastrous for a libertarian conception of free will. The experience of other options open to us, so naturally accompanying deliberation of our actions, could prove to be a deceptive illusion. We would not be free agents, responsible for our actions, only ill-informed perceivers of what was long construed

out of our conscious control. Hypothesis 1 would be no more a hypothesis, but an established fact.

We have little to say about the Hypothesis 2, apart from stating an obvious: no important discovery, linking the domain of quantum mechanics to the volitional acts has been made and we are quite skeptical there will be any advance accomplished in the foreseeable future.

However, we do not want to end up in a full agreement with original Hypothesis 1. It is because, just like Searle, we cherish the notion of freedom and its position within the concept of humanity and human society. He speaks for us when he writes: "It seems to me we find the psychological experience of freedom so compelling that it would be absolutely astounding if it turned out that at the psychological level it was a massive illusion, that all our behaviour was psychologically compulsive" (Searle 2001, 496). A possibility to choose, an existence of an array of available options and our position of action initiators appear to us so profoundly rooted in our self-conception that we feel its loss would undermine the very essence of what makes us the kind of beings we are. Instead of postulating an irreducible self or searching for gaps in the fundamental build-up of the world, we suggest a simpler route. It is one that takes first person ontology seriously and clarifies the framework of the third-person ontology in order to dissolve an apparent conflict between them. We suggest to rethink the notion of determinism and handle it not as a given fact, but rather as an unfounded philosophical myth that looks ever-present while in fact it is nowhere to be found. This position traces its roots to the observation of Russell: "All philosophers, of every school, imagine that causation is one of the fundamental axioms or postulates of science, yet, oddly enough, in advanced sciences such as gravitational astronomy, the word "cause" never appears" (Russell 1912, 1). Indeed, there are laws, equations, particles, fields and who knows what across scientific fields, but no mention of causes, not to speak of determinism. It is possible that attempts to establish a solution to determinism are fighting a straw man. Maybe there is no evidence for determinism, because there is no evidence for omnipresence of causes in the natural world. If science as our best epistemic practice operates without them, why should a naturalist be worried?

Notice that the view has the virtue of respecting first-person and third-person ontologies, because it is faithful to experiences of conscious gaps and at the same time is open to challenges from natural

science that might undermine volition. This is because even if science comes up with results like those reported above, it only undermines our sense of agency in a very limited domain. The artificial conditions of the experimental set-up, individual or social differences and various other factors have to be excluded for the experiment to have the kind of over-reaching consequences that many would like to see. And even if it is eventually extended to a wider range of phenomena, there is no reason to believe that what it demonstrates are wider gaps in consciousness than we thought we had, memory lapses, post-dictions and other purely psychological phenomena, that, as Searle repeatedly argues, are realized by the grey matter. Yet it is by no means indicative of a wider deterministic nature of the world that would make us puppets, enclosed in our phenomenal minds without real effects in the world. Causal chains on this picture can start anywhere and causal closure is just an assumption that needs to be first firmly established and only then taken seriously. Let us transfer the burden of proof for such an uneasy endeavor on those who believe the truth of determinism. If they are ever successful, then we will have to worry about how to make claims of freedom and determinism compatible. For now, as well for a distant future, we see no reason to worry about impossibility of being free.[1]

Institute of Philosophy
The Academy of Sciences of the Czech Republic, v. v. i.
Jilská 1
110 00 Praha 1
Czech Republic
hvorecky@flu.cas.cz

References

Bode et al. (2011): Tracking the unconscious generation of free decisions using ultra-high field fMRI. PloS One, 6, 1-13.

Russell, B. (1912): On the Notion of Cause. *Proceedings of the Aristotelian Society* 13, 1-26.

Searle, J. (2001): Free Will as a Problem in Neurobiology. *Philosophy* 76, 491-514.

Soon et al. (2008): Unconscious determinants of free decision in the human brain. *Nature Neuroscience* 11, 5, 543-545.

Thoughtful Brutes

Tomáš Hříbek

The Academy of Sciences of the Czech Republic, Prague

Abstract: Donald Davidson and John Searle famously differ, among other things, on the issue of animal thoughts. Davidson seems to be a latter-day Cartesian, denying any propositional thought to subhuman animals, while Searle seems to follow Hume in claiming that if we have thoughts, then animals do, too. Davidson's argument centers on the idea that language is necessary for thought, which Searle rejects. The paper argues two things. Firstly, Searle eventually argues that much of a more complex thought does depend on language, which reduces a distance between himself and Davidson. Secondly, some of Davidson's suggestions are promising – in particular the idea that we may lack a vocabulary to capture the contents of animal thoughts. Based on this insight, one might, *pace* Davidson, grant thoughts to animals. However, this does not mean, *pace* Searle, that it should be possible to construe even the simplest of such thoughts as propositional. Perhaps we need to move beyond Davidson *and* Searle by developing a theory of non-propositional thought for animals.

Keywords: Donald Davidson, John Searle, animal minds, semantic holism, *de re/de dicto*, triangulation, non-propositional thought.

1

Do animals other than humans think? Major philosophers of the past, such as Descartes and Hume, expressed early contrasting views concerning this issue. As is well known, Descartes sharply denied any mentality—not just beliefs, but also consciousness—to animals, while Hume opined that "no truth appears to me more evident, than that beasts are endow'd with thought and reason as well as men" (Hume 1978, 176). Descartes' position was motivated by theological con-

cerns—if animals entertained thoughts, we would have to worry about the fate of their souls, as thinking is the activity of the immortal soul—but his main argument can be stated in abstraction from theology. In a nutshell, animals do not think, because they do not speak—where language was, for Descartes, evidence of the presence of thought. Hume argued for the opposite conclusion on the basis of a similarity between human and animal behavior: given that we know by introspection that human behavior is accompanied by "ideas", the closely resembling behavior of brutes is likely to be accompanied by such inner episodes, too.

Two developments characterize the current debate on animal minds. First, unlike the early moderns for whom the issue of animal mentality remained peripheral, contemporary philosophers see it as central. Second, the current debate decidedly favors Hume's position over Descartes'. Most contemporary thinkers feel that if *we* possess minds, then other animals are bound to possess at least rudiments of mentality as well. To be sure, Darwin rather than Hume is a direct influence on the contemporary philosophy of animal minds. Whereas Hume offered a mere argument by analogy between human and animal behavior, Darwin supplied a testable hypothesis about the continuity between humans and the rest of creation. From the Darwinian point of view, provided that other creatures are our evolutionary kin, it is absurd to believe that thought has not emerged gradually, like other traits, and that it does not exist in simpler forms. As Darwin put it in *The Descent of Man*:

> If no organic being excepting man had possessed any mental power, or if his powers had been of a wholly different nature from those of the lower animals, then we should never have been able to convince ourselves that our high faculties had been gradually developed. But it can be shown that there is a much wider interval in mental power between one of the lowest fishes, as a lamprey or lancelet, and one of the higher apes, than between an ape and a man; yet this interval is filled up by numberless gradations. (Darwin 1871, 44)

It is the conviction that philosophical theorizing about the nature of thought must respect our best empirical knowledge about a common origin of species that motivates the moving of the topic of animal mentality to the forefront of current debate. Professor Searle contributed to this reorientation with his paper "Animal Minds" (see Searle 1994). Much of it is taken up by Searle's critique of various arguments by Don-

ald Davidson, all of which assume that the possession of a language is somehow necessary for thought. Early on in the paper, Searle voices a deep mistrust of any arguments that ignore empirically attested similarities between humans and other animals. The late twentieth-century debate between Searle and Davidson thus seems to repeat the structure of the early modern disagreement between Hume and Descartes. To be sure, there are important differences. First, unlike Descartes, Davidson does not view language merely as a sure sign, but rather as constitutive, of thought. Second, unlike Hume, Searle does not search for relevant similarities between humans and animals that justify a belief in animal minds at the level of outward behavior, but rather in neurobiology. Third, unlike with Descartes and Hume, there is no fundamental disagreement between Davidson and Searle in metaphysics, as they both assume that the world is fundamentally composed of physical entities.[1] And yet I think we can say with a bit of license that Searle plays a latter-day Hume to Davidson's Descartes. Searle continues the naturalistic program in philosophy, of which Hume was a founder, which is opposed to any tendency to deny certain phenomena in the name of an a priori theory. This is what I see as the key critical insight of Professor Searle's in his debate with Professor Davidson: if a philosophical theory of thought is so demanding that it denies thought to nonhuman animals, of whom both common sense and science assumes otherwise, then so much the worse for the philosophical theory.

However, I do not wish to simply conclude that Davidson was wrong to deny thought to animals, while Searle is right to grant it to them. For one thing, even though Searle does not take language as necessary for thought, he eventually argues that much of a more complex thought does depend on language, which reduces a distance between himself and Davidson. Secondly, I believe Professor Davidson's arguments can be mined for insights that can help us to get ahead in the philosophy of animal minds. I am thinking, in particular, of the suggestion that we may lack a vocabulary to capture the contents of animal thoughts. This is what I am going to argue, based on this insight: We must, *pace* Davidson, grant thoughts to animals. However, this does

---

[1]   The doctrine that the world is ultimately physical is traditionally called "materialism". Davidson is one of many recent philosophers who adopted the doctrine and the label. Professor Searle eschews the label, but the discussion of his reasons for this terminological decision is beyond the scope of my paper.

not mean, *pace* Searle, that it should be possible to construe even the simplest of such thoughts as propositional. Perhaps we need a theory of non-propositional thought for animals, although I cannot hope to fully develop it here. But let me begin by reviewing Searle's critique of Davidson's arguments.

2

Professor Searle identifies two separate arguments against animal thought that he ascribes to Donald Davidson. I shall revisit them in turn, adding my own comments to Searle's critique.

The first of Davidson's arguments turns on the idea that in ascribing thoughts to each other, we make fine discriminations with respect to their contents that seem impossible in the absence of language. As animals don't have language, there seems to be no way to pin down what thought they might be having in mind. As Davidson puts it in his paper "Thought and Talk":

> The dog, we say, knows that its master is home. But does it know that Mr. Smith (who is his master), or that the president of the bank (who is that same master), is home? We have no real idea how to settle, or make sense of, these questions. (Davidson 1984 [1974], 163)

In response, Professor Searle claims that the argument assumes a verificationist premise that unless it's possible to determine the propositional content of a thought ascribed to the dog in Davidson's example, it makes no sense to ascribe it in the first place. Searle rejects the premise and goes on:

> Even if we assume that there is no fact of the matter as to which is the correct translation of the dog's mental representations into our vocabulary; that, by itself does not show that the dog does not have any mental representations, any beliefs and desires, that we are trying to translate. (Searle 2002 [1994], 66)

I think Professor Searle is right that Davidson's first argument is inconclusive. But I wish to add two points on Davidson's behalf. For one thing, he is not committed to verificationism. Indeed, he explicitly disowns it in a later paper, "Rational Animals" (see Davidson 1982). He says that he assumes that "an observer can under favorable circumstances tell what beliefs, desires and intentions an agent has" (Davidson 2001 [1982], 99). But, he adds, "[m]erely to claim that an observer can

under favorable conditions tell what someone else is thinking is not to embrace verificationism, even with respect to thought" (Davidson 2001 [1982], 99). More importantly, Davidson himself acknowledges, also in "Rational Animals," that his arguments does not establish the strong conclusion that there can be no thought in the absence of a language, but at best a weaker thesis that "there probably can't be much thought without language" (Davidson 2001 [1982], 101). This is how Davidson seems to make a room for primitive animal thought: Even though we cannot capture the exact way a speechless animal, such as a dog, thinks of some object, we can come up with some description of the object that the dog could pick out. To use a bit of jargon, the fact that we may not be justified in ascribing *de dicto* beliefs to nonlinguistic animals does not rule out the possibility that they lack beliefs *de re*. However, Davidson adds that ascribing any single belief presupposes ascribing indefinitely many more. And, as we have no way to tell whether a speechless animal has any of these additional beliefs, especially more complex ones, we are not on a very solid ground even with the ascription of the very simple beliefs. But, as said before, the first argument does not rule them out completely.

However, Professor Davidson attempts to prove the strong thesis that animals lack even *de re* thoughts in another argument. That is, he questions the possibility of identifying any objects of purported animal thoughts, not just our ability to capturing the ways animals might conceive of such objects. It is this argument that qualifies Davidson as a Cartesian, despite his materialist metaphysic. An early version of this argument can be found in the last few pages of his "Thought and Talk," and this is the version critically analyzed by Professor Searle. He summarizes the argument in three steps. At a first step, Davidson repeats the holistic assumption that we've already seen in the previous argument, namely that in order to be ascribable any particular thought, an animal must possess a whole set of beliefs. At a second step – which is crucial – Davidson claims that in order to have beliefs, an animal must have the concept of belief. Third, in order to have the concept of belief, one must have a language. Yet animals do not have any language; therefore, they do not think.

In his commentary, Professor Searle does not question the first premise, i.e. the holistic assumption that a thought can be ascribed only against the background of a whole lot of beliefs. After all, Searle argued for a similar kind of semantic holism in his own work—e.g., in chapter

8 of his book *The Rediscovery of the Mind* (1992), in which the idea of holism appears under the label "Network". To be sure, not everybody accepts holism. For example, Jerry Fodor questions the idea for a variety of reasons, one of which is that he thinks it is inhospitable to the possibility of simple animal minds. But I shall leave that route of criticism of Davidson's argument aside. Instead, Professor Searle concentrates on Davidson's second premise, which expresses the assumption that a minded creature must possess the concept of belief. Davidson's rationale for making this assumption was that the possession of belief presupposes an ability to distinguish between true and false beliefs, for which the possession of metalinguistic semantic predicates such as "belief" is mandatory. Searle questions this strong assumption. He says:

> I agree [that] … having an intentional state requires the capacity to discriminate conditions which satisfy from those that do not satisfy the intentional state … [But]… I see no reason at all to suppose that this necessarily requires a language, and even the most casual observation of animals suggests that they typically discriminate the satisfaction from the frustration of their intentional states, and they do this without a language. (Searle 2002 [1994], 67)

Searle further argues that we may miss how this discrimination naturally works, if we forget that perception and action—not just belief—are also forms of intentionality. But perception fixes belief and belief determines action. A dog believes that a cat is up a tree because he saw and smelled the cat running up the tree, which leads him to chase the cat and bark up the tree. And so on.

Again, I agree with Professor Searle that Davidson's second, more radical argument against animal thought is also inconclusive. Yet I should like to leave his point about perception as a form of intentionality aside for a moment; I shall return to it at the end of this section. Instead, I wish to consider a bit more complex form of the same argument against *de re* animal belief which Davidson elaborated in "Rational Animals."

In this later version, Davidson attempts to gather more support for his controversial premises, in particular for his claim that the possession of a belief requires the concept of belief. Davidson tries to support this claim by an additional argument which goes roughly as follows (cf. Davidson 2001 [1982], 104): In order to have beliefs, one must be capable of being surprised. To be surprised means to realize that one,

or more, of one's beliefs were incorrect. However, that means that one possesses the concept of belief. Hence follows the conclusion that in order to have beliefs, one has to have the concept of belief. Obviously, the most questionable premise in this supporting argument is the first one. I surmise that Professor Searle, had he considered this suggestion, would have objected to it as a piece of speculative psychology. It seems preposterous to attempt to decide a priori whether or not a minded creature must be capable of surprise. Yet perhaps Davidson could have arrived at the conclusion that a creature with beliefs needs to possess the concept of belief by an alternative route. Let us say that someone is a thinking creature if she possesses lots of attitudes with respect to her present circumstances as well as the future. However, that seems to presuppose a capacity to realize that, for instance, one is carrying out one's plans and that they go well or badly. But this seems to presuppose the concept of belief. So we can get at Davidson's desired conclusion without a priori assumptions about the capacity for surprise. Yet a critic might object that the concept of a thinker that is presented here is too high-brow to be applicable to non-human animals, but I think this conclusion is premature and should await empirical testing. At any rate, Davidson's premise that having beliefs presupposes the concept of belief is not hopeless, and it is an open question whether or not it is applicable to non-human animals.[2]

However, Davidson further needs to support the third premise of his second argument, namely the claim that one can have the concept of belief only if one has a language. Davidson's argument here is dense in the extreme, involving such fundamental concepts as truth, objectivity and communication. Perhaps it can be reconstructed as follows (cf. Davidson 2001 [1982], 104-105): In order to have the concept of belief, one must have the concept of truth – i.e., one must be capable of contrasting between what is believed and what is the case. One could come to possess the concept of truth only if one were involved in communication with another creature, since the concept of truth, finding out how things are objectively, would play a crucial role in interpreting the other creature. Hence in order to have the concept of belief, one must be in communication with another. Now, as communication is conducted

---

[2]   Cf. Allen – Bekoff (1997) for evidence that members of various species other than human are capable of feats such as deception and self-recognition that presuppose second- if not higher-order beliefs.

in a language, in order to have the concept of belief, one must communicate with others by means of a language.

This seems to complete Davidson's defense of steps two and three of his second argument against animal thought. However, at the very end of "Rational Animals" he describes a model context in which the concepts of truth and objectivity arise. It involves an early occurrence of the idea of triangulation that Davidson went on developing in his last papers. The passage reads:

> If I were bolted to the earth, I would have no way of determining the distance from me of many objects. I would only know they were on some line drawn from me toward them. I might interact successfully with objects, but I could have no way of giving content to the question where they were. Not being bolted down, I am free to triangulate. Our sense of objectivity is the consequence of another sort of triangulation, one that requires two creatures. Each interacts with an object, but what gives each the concept of the way things are objectively is the base line formed between the creatures by language. The fact that they share a concept of truth alone makes sense of the claim that they have beliefs, that they are able to assign objects a place in the public world. (Davidson 2001 [1982], 105)

The idea seems to be that two individuals respond to the same object and to each other, thus forming a triangle, where the base line is their communication, and this is the only way how the creatures could come to possess the concept of objective truth. However, in some of his later papers it seems that Davidson puts the idea of triangulation to a somewhat different use; it serves to provide an alternative route to the conclusion that language is necessary for thought. Davidson seems to argue that in the absence of actually communicating with another creature, one could not come by a determinate object of thought:

> It takes two points of view to give a location to the cause of a thought, and thus to define its content. We may think of it as a form of triangulation: each of two people is reacting differentially to sensory stimuli streaming in from a certain direction. Projecting the incoming lines outward, the common cause is at their intersection. If the two people now note each others' reactions (in the case of language, verbal reactions), each can correlate these observed reactions with his or her stimuli from the world. A common cause has been determined. The triangle which gives content to thought and speech

is complete. But it takes two to triangulate. (Davidson 2001 [1991], 212-13)

What we seem to be getting in this last passage is a fantastically strong claim: there is no thought, no cognitive content, outside of an *actual* communication with another speaker. By comparison, the previous arguments involving the concepts of truth and objectivity seemed to be supporting a weaker thesis that such concepts only arise in the context of a communication.

Davidson's arguments against *de re* animal thought are bound to be unsound, since they end up ruling out not just the possibility of animal thought, but also of solitary human thought. On the latter point, one wonders why it should be impossible to triangulate with oneself, so to speak – say, by noticing gradual changes in the world and checking on the external objects from different points of view. It is pretty clear that Davidson's absurd view is a consequence of an impossibly demanding theory of thought which sees thought and belief as constitutively dependent on language. However, such a high-minded conception of thought seems especially hopeless especially in view of the situation in behavioral sciences of the last few decades. These disciplines have witnessed a "cognitive turn," which means that it has become customary to ascribe rather sophisticated cognitive capacities to languageless creatures. Thus, in the thriving field of cognitive ethology, we find research programs that start from the assumption that animals have beliefs and desires and act on them. However, a similar approach is taken by contemporary developmental psychologists who ascribe to prelinguistic infants a "theory of the mind." And the important fact is that these disciplines get integrated in a single, broadly Darwinian, naturalistic-cognitivist paradigm. A philosophical theory that contravenes these developments looks like a relic from a prescientific era.

I take it that Professor Searle sees Davidson's failure to grant cognitive lives to animals other than humans as a *reductio ad absurdum* of his whole philosophical project. Searle's easy accommodation of non-human thinkers should then be seen as evidence of a superiority of the reverse methodology, which he systematically developed in an earlier book, *Intentionality* (1983). According to this methodology, the mind has a priority over language:

> From an evolutionary point of view, just as there is an order of priority in the development of other biological processes, so there is an order of priority in the development of Intentional phenomena.

> In this development, language and meaning, at least in the sense in which humans have language and meaning, comes very late. Many species other than humans have sensory perception and intentional action, and several species, certainly the primates, have beliefs, desires, and intentions, but very few species, perhaps only humans, have the peculiar but also biologically based form of Intentionality we associate with language and meaning. (Searle 1983, 160)

I think that Searle's developmental approach to intentionality can be exploited in response to Davidson's thesis discussed above that outside of triangulation, there is no way to fix the contents of thoughts. For this purpose, we should revisit Searle's idea that perception is a form of intentionality, and that animals are naturally attributed perceptual beliefs. Even granted, for the sake of an argument, that there are linguistically mediated forms of thought such that an interpreter needs to be involved in them, it is a natural consequence of Searle's evolutionary approach that perceptual representation is older in the order of phylogeny, so that its content must be determined independently of an interpreter. This sort of content gets determined in terms of what an animal can discriminate and how it is capable of using these discriminations in the ways it navigates through its environment.

Furthermore, Searle's developmental methodology makes it possible to preserve the rational core of Davidson's denial of animal thought, namely the idea that crucial kinds of thought entertained by humans are indeed unavailable to beasts that lack a language. For example, Searle argues that, without a language, animals cannot entertain metalinguistic thoughts (such as that "eat" is a transitive verb); they cannot think of institutional facts of which language is constitutive (such as that this piece of paper is a legal bill); they cannot represent to themselves facts so remote in space and time that they are unavailable without language (such as the facts that obtained in the past); and they cannot think logically complex facts, such as subjunctive facts; etc. And in a recent book, *Rationality in Action* (2001), Professor Searle draws further consequences from the fact that speechless animals cannot perform certain important speech acts, such as asserting. Due to this inability, they cannot have desire-independent reasons for action. Consequently, non-linguistic animals act only in order to satisfy some non-rational desire or other. If such acts as courage or loyalty involve desire-independent reasons, it follows that a non-linguistic animal cannot ever be courageous or loyal.

3

And yet I am going to conclude by suggesting that what we might need to do in the philosophy of animal minds is to move beyond both Davidson *and* Searle. Let me explain. I have been critiquing Davidson for his anti-naturalistic tendencies. Yet he was as most other recent philosophers indebted to Quine and wished that philosophy was respectful of, if not reducible to, sciences. We might be helped to understand better Davidson's point of view, if we turn to one of his late papers, "The Emergence of Thought" (1999). In it, he clearly accepts a developmental or evolutionary point of view, but he argues that there is a conceptual difficulty in describing transitions from one level of development to the next, if each of these levels is characterized in terms of different concepts:

> In both the evolution of thought in the history of mankind, and the evolution of thought in an individual, there is a stage at which there is no thought followed by a subsequent stage at which there is thought. To describe the emergence of thought would be to describe the process which leads from the first to the second of these stages. What we lack is a satisfactory vocabulary for describing the intermediate steps. (Davidson 2001 [1999], 127)

What this passage seems to suggest is that Davidson does not wish to deny what today must be obvious to every scientifically educated person, namely that the world has developed from dead matter through mere sensation to thought and language. Rather, Davidson cautions that we do not have a conceptual wherewithal to describe the emergence of new levels. We can interpret this as a new version of his earlier argument against *de dicto* animal thought. It is compatible with what Davidson is saying that our hominid ancestors had cognitive lives in some sense, even prior to the emergence language. It is possible that prelinguistic infants think. And it is possible that members of other species think. It is just that we have no means of identifying the contents of thoughts other than sentences of some natural language or other. As with the earlier argument, this does not mean that non-linguistic creatures have no minds; it's just that we are not justified in granting thoughts under these circumstances.

Let is turn now to Professor Searle's theory to see whether it fares better when it comes to the problem of emergence of thought. We saw that, for Searle, there is a prelinguistic level of intentionality which

confers content onto language. The prelinguistic level of intentionality is supposed to be a natural feature of certain complex biological systems. This might be difficult to make sense of, but I shall leave this issue aside. The problem that I wish to point out is rather the opposite of the one we saw earlier in Davidson. Recall that Davidson had a trouble to make sense of thought *sans* language. This seems easily solved for Searle, for whom thought has a genealogical as well as logical priority over language. According to the theory proposed in *Intentionality*, utterances in a natural language get their meaning from prelinguistic intentions. The more complex the utterances, however, the more difficult it seems to accept that prelinguistic intentions could have complex propositional structures of sentences of a natural language. So while on Searle's account we can speak with no qualification of animals' thoughts, it seems puzzling that there could be prelinguistic intentions of the required complexity in adult humans to confer the required contents onto their speech. Interestingly, Professor Searle seems to concede that most forms of intentionality in mature humans are linguistic, which would seem like a concession to Davidson (cf. Searle 1991, 94). But he still insists that the meanings of language can be explained in terms of the intentionality of the mind. Where does the complexity of the latter come from?[3]

I announced in the beginning that I was going to suggest a way out, an alternative to both Davidson and Searle on animal thought. I can be only brief now. Davidson turned out to be a sort of eliminativist with respect to animal thought, while Searle is a realist. Both theories, however, seem to share the assumption that thought is inherently propositional. Davidson argued that thoughts inherit their propositional structure from a language. Searle maintains that this structure is conferred by prelinguistic intentions. I shall leave aside which of the two theories makes better sense of the adult human thought. However, what we see in current cognitive ethology and related disciplines is perhaps evidence of forms of thought that are not necessarily propositional in character. While there is a broad consensus as to the notion that nonhuman

_____

[3] One suggestion, popularized by Jerry Fodor, is that the underlying structure is indeed propositional; but it does not come from a natural language. Instead, it comes from the so-called language of thought. Searle, however, emphatically rejects this proposal, as it is connected with the whole computational approach to intelligence that he demolished in his famous critique of artificial intelligence (cf. Searle 1980).

animals are thinkers, their thinking is perhaps better characterizable as an exercise in imagination; or in terms of achieving a certain goal, even though the process leading up to that goal cannot be expressed in terms of any explicit propositions.

Institute of Philosophy
The Academy of Sciences of the Czech Republic, *v. v. i.*
Jilská 1
110 00 Praha 1
Czech Republic
tomas_hribek@hotmail.com

References

ALLEN, C. – Bekoff, M. (1997): *Species of Mind*. Cambridge, Mass.: The MIT Press.

DARWIN, C. (1874). *The Descent of Man*. London: J. Murray.

DAVIDSON, D. (1974): Thought and Talk. In: Davidson, D. (1984): *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press, 155-170.

DAVIDSON, D. (1984): *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press, 1984.

DAVIDSON, D. (1985): Rational Animals. In: Davidson, D. (2001): *Subjective, Intersubjective, Objective*. Oxford: Oxford University Press, 95-105.

DAVIDSON, D. (1991): Three Varieties of Knowledge. In: Davidson, D. (2001): *Subjective, Intersubjective, Objective*. Oxford: Oxford University Press, 205-220.

DAVIDSON, D. (1999): The Emergence of Thought. In: Davidson, D. (2001): *Subjective, Intersubjective, Objective*. Oxford: Oxford University Press, 123-134.

DAVIDSON, D. (2001): *Subjective, Intersubjective, Objective*. Oxford: Oxford University Press.

SEARLE, J. (1980): Minds, Brains, and Programs. *The Behavioral and Brain Sciences*, vol. 3.

SEARLE, J. (1983): *Intentionality*. Cambridge: Cambridge University Press.

SEARLE, J. (1992): *The Rediscovery of the Mind*. Cambridge, Mass.: The MIT Press.

SEARLE, J. (1991): Responses to Critics. In: Lepore, E. – Van Gulick, R. (eds.): *John Searle and His Critics*. Oxford: Blackwell.

SEARLE, J. (1994): Animal Minds. In: Searle, J.(2002): *Consciousness and Language*. Cambridge: Cambridge University Press, 61-76.

SEARLE, J. (2002): *Consciousness and Language*. Cambridge: Cambridge University Press.

SEARLE, J. (2004): *Rationality in Action*. Cambridge, Mass.: The MIT Press.

Intentionality and What We Can Learn about It from Searle's Theory of Institutions

Pavla Toráčová

Charles University, Prague

Abstract: Searle's theory of institutions is based on the insight that institutional facts are created in intentionality, and it consists in the logical analysis of the intentional performance in which the institutional facts are created. The aim of this paper is to relate Searle's account of intentionality as creating institutional facts to his general account of intentionality elaborated in his book *Intentionality. An Essay in the Philosophy of Mind*. I come with the claim that the imposition of status function, that characterizes the intentional performance in which institutional facts are created, consists in double prescription of conditions of satisfaction, where ones of them are related to our goals and interests while the other ones are independent of them. I suggest that this holds true for intentionality in general.

Key words: conditions of satisfaction, institution, intentionality, status function.

Intentionality – a property of mental states that is to be "about something" – is one of the most mysterious phenomena. We have no idea how this phenomenon comes into being, nor do we know what exactly it consists in. How can there be processes in the world of nature – most probably some processes in the brains of animals – that are "about something"? How could it happen that the world, at one point in its history, became aware of itself? Natural processes of some special kind became "about" another natural processes, and later also about many other things – about things that neither ever existed nor ever will exist, or about things whose existence is created simply by the fact that they

are the objects of intentional states. But what is it exactly to be about something or to be intentionally related to something?

John Searle addressed this problem in his book *Intentionality. An Essay in Philosophy of Mind* (see Searle 1983). In this book, he not only analyzed the logical properties of intentional states, but also insisted on strictly distinguishing between a *logical* and an *ontological* way of talking – the misunderstanding of which is the source of much confusion in philosophy. The clarity in this matter is the key to understanding Searle's theory of intentionality – and perhaps also to understanding Searle's philosophy as such.

Searle's account laid the foundation for a general theory of intentionality describing the essential logical properties that characterize the different types of intentional states. His book represents a breakthrough in the study of intentionality in the philosophy of mind; the reader, however, will want to know more. Thanks to Searle, we know that the intentional state is characterized by intentional content that often determines the conditions of satisfaction, and that this intentional content is in a certain *psychological mode* which – in some cases – determines its *direction of fit*. But what exactly does *to have intentional content* and *to determine the conditions of satisfaction* consist of? It seems that the ability *to have the content that determines the conditions of satisfaction* is crucial for intentionality, but there is no account that works out these "properties" – neither in Searle's philosophy nor in contemporary philosophy of mind.

How is Searle's theory of institutions to help us in these questions? The theory is based on the insight that institutions and institutional facts as such are created in intentionality. They are facts only in so far as people take them to be facts in their beliefs, desires, intentions, expectations and other intentional states. Intentionality then represents a performance that creates the object that it is "about". Thus it would seem a promising endeavor to study what Searle's account of this performance says about intentionality in general.

Before we proceed to do this, we should review what Searle says about intentionality in his book on the subject. He describes four categories that characterize intentional states: *intentional content*, *psychological mode*, *direction of fit*, and *conditions of satisfaction*. The intentional content is not something that the agent would be related to, or what he or she would *use* as a "mediator" to relate to the intentional object, it is rather the state itself. In this way, every intentional state has certain in-

tentional content (see Searle 1983, 6). The intentional content is given in certain psychological mode, such as belief, desire, fear, hope, perceptual experience, etc. The psychological mode determines whether and what kind of direction of fit the intentional state has. Some intentional states – belief or perceptual experience may be examples – are supposed to match the world: they have a *mind-to-world* direction of fit. Some other states, like intentions or desire, are supposed to bring about changes in the world so that the world matches their content – they have a *world-to-mind* direction of fit (cf. Searle 1983, 8). In many types of the intentional states, the intentional content can be expressed by a proposition. Where the intentional content is propositional and the state has a direction of fit, the intentional content determines its *conditions of satisfaction*. Conditions of satisfaction are those conditions which must be obtained if the state is to be satisfied (see Searle 1983, 12-13).

Every intentional state has an intentional content in a certain psychological mode. In some intentional states – actually in the most important types of them such as intentions, perceptual experiences, beliefs, desires – the intentional content determines the conditions of satisfaction of that state, and the psychological mode determines its direction of fit. It seems to me that we could see these four notions that characterize intentionality as actually being just two notions: (1) the intentional content that takes the form of the conditions of satisfaction for some states, and (2) the psychological mode that, for some states, determines the direction of fit.

At least two more things should be said about the conditions of satisfaction: they are always represented under some aspects and, in some intentional states (perceptual experiences and intentions in action are basic examples of them), they have a *self-referentiality* (a prescription of a causal relation) incorporated in them.

This account of intentionality, according to Searle, says nothing about the *ontology* of intentional states; it instead addresses their *logical* properties.

> If the question 'What is a belief really?' is taken to mean: what is a belief *qua belief*?, then the answer has to be given, at least in part, in terms of the logical properties of belief: a belief is a propositional content in a certain psychological mode, its mode determines a mind-to-world direction of fit, and its propositional content determines a set of conditions of satisfaction. Intentional states have to be characterized in Intentional terms if we are not to lose sight of their

intrinsic Intentionality. But if the question is 'What is the mode of existence of beliefs and other Intentional states?' then from everything we currently know about how the world works the answer is: Intentional states are both caused by and realized in the structure of the brain. (Searle 1983, 15)

Keeping this in mind will help us to not make any of the common philosophical mistakes and confusions: the intentional contents are not "mysterious entities", it is the way in which we characterize the intentional states if we are concerned with their intentionality. Neither are they images in our heads, nor do they contain some sense data. The contents are prescriptions of the conditions of satisfaction that are satisfied by the intentional objects but they don't have the properties that they ascribe to the objects – e.g. the visual experience of yellow is not yellow.

The *intentional content* or the *conditions of satisfaction* is a name for a logical property whereas the *intentional object* names an ordinary object: "an Intentional object is just an object like any other; it has no peculiar ontological status at all" (Searle 1983, 16).

There are some other important features about intentional states. For example, the necessity for the intentional states to be a part of a *Network* of other intentional states and to stand against a *Background* of practices and preintentional assumptions (cf. Searle 1983, 19-21).

We can summarize: In order for a state to be intentional, it has to have an intentional content with which it relates to its object; the manner of this relation being determined by the psychological mode of the state. In the cases of the intentional states that can be called *paradigmatic*, like beliefs, desires, perceptual experiences and intentions in action, the relation of the state to its object can be characterized as a "fit", in which the responsibility for that fit can rest either on the intentional state (mind-to-world direction of fit), or on the world (world-to-mind direction of fit). In these cases, the intentional contents determine their conditions of satisfaction. As a result of having all these properties, the state can have its intentional object, or it can be *of* or *about* an object (providing the conditions of satisfaction are satisfied). In the paradigmatic cases, the intentional object is a real object in the world, or, more precisely, it is a state of affairs in the world (which is reflected in the propositional form of their intentional content).

It is the ability to have a content that prescribes conditions of satisfaction with a certain direction of fit that is the defining mark of inten-

tionality. It could seem from Searle's account that the content alone is not enough for a definition of intentionality, and that a psychological mode has to be added in order that the state is intentional. But the function of the psychological mode is to determine the direction of fit, and the direction of fit is part of the conditions of satisfaction prescribed by the content – quite similarly as self-referentiality is a part of the conditions of satisfaction in the case of the perceptual experience or the intention in action.

If intentionality is defined as an ability to have a content that prescribes conditions of satisfaction with a certain direction of fit, we might then want to know more about the content. What is its structure or composition? What exactly does it consists of to have this kind of content?

There should be something in the intentional content that reflects the ability to prescribe the conditions of satisfaction and to have the direction of fit. We should be able to see what makes the content be a prescription of the conditions of satisfaction with certain direction of fit. Further, there should be something in the intentional content that reflects the logical relations and connections between the types of intentional states. There are at least two kind of relations between intentional states: first, the systematic ones, as every intentional state has to be a part of a network of intentional states – a belief has to be related to other beliefs, intentions, perceptions, etc. Secondly, there are some "hierarchical" relations to be found: some intentional states are more primordial or primary than others, e.g., the perceptual experiences and intentions in action are, as Searle says, "biologically primary forms of intentionality" because they have intentional causation in their conditions of satisfaction, and beliefs and desires appear to be rather "etiolated forms of more primordial experiences in perceiving and doing" from which the intentional causation has been "bleached out" (Searle 1983, 36). Furthermore, the perceptual experiences and the intentions in action are also primary to memories and prior intentions which also involve some kind of intentional causation but presuppose perceptions and intentional actions and build on them. The primacy is not only "biological", but also logical. A further analysis of the intentional content should shed light on all these logical relations.

At this point, it is necessary to clarify what kind of analysis we are calling for. Searle denies that the analysis of the *formal* structure of intentional states is a relevant method for the investigation of intentional

states *qua* intentional. Intentionality is defined by the content, not by some formal structure, as artificial intelligence and cognitive psychology often assume (see Searle 1983, 12). We cannot investigate semantics through the investigation of syntax. This opinion of Searle's is something that we agree with. If we call for further analysis of the intentional content, and perhaps for an exposition of its general "structure", we have in mind the investigation of the structural elements of the *content*, not of the syntax or some other formal structure. The generality here does not imply formality.

Searle emphasizes that the intentional content always determines the conditions of satisfaction *under certain aspects*. The aspectual character implies the first person perspective that should be taken into account if we want to investigate the intentional content. But the first person perspective is a concept that is philosophically ambiguous: for many philosophers, it implies something like a "private character" of the content, a kind of epistemic privacy in the sense that Wittgenstein was worried about. However, this implication is by no means necessary. While it is true that the agent always experiences the world from his or her point of view and that other people experience the same world from their point of view, it doesn't follow that the perspectival mode of experience is essentially private or incommunicable. And if we want to investigate the intentional content *qua intentional*, we have to respect its aspectual, and therefore also first-personal character. We seek to describe the general features of the content, which is to say we seek to find some general features in what the agent "comprehends" when she lives in the flux of her conscious intentional life. In order to do it, we don't need to get into her "epistemically private zone" (and to ponder whether and how it is possible). All that we need to do is to keep a clear understanding of which properties are relevant to our investigation and which are not relevant.

I have mentioned the *conscious* form of the intentional states. But isn't this a complication of the investigation? So far, we have been dealing with intentional states and there is no necessity for those states to be conscious – we have, for example, many beliefs that have never been brought to consciousness. I would like to say that the notion of consciousness seems to be implied by the first-personal perspective as such. How else would the intentional states acquire their first-personal character if not through the fact that they can be, at least potentially, conscious? Also, as Searle says in his book *The Rediscovery of the Mind*:

"we do not have a clear notion of unconscious mental states. … *The notion of an unconscious mental state implies accessibility to consciousness.* We have no notion of the unconscious except as that which is potentially conscious" (Searle 1992, 152).

Let's us now proceed to explore what Searle says about the intentional content, the conditions of satisfaction and the direction of fit in his account of social and institutional reality. It is true that, explicitly at least, he doesn't say anything about them. However, his account of social and institutional reality is essentially based on his investigation of intentionality. Institutional reality comes to existence in the performance of intentionality by a human community and it continues to exist only as far as the intentionality maintains it in existence. This is the basic insight that determines Searle's point of departure. In the beginning of his book *The Construction of Social Reality* Searle writes:

> …there are portions of the real world, objective facts in the world, that are only facts by human agreement. In a sense there are things that exist only because we believe them to exist. (Searle 1995, 1)

Searle calls these facts *institutional facts* and as examples of them he states:

> I am thinking of such facts as that I am a citizen of United States, that the piece of paper in my pocket is a five dollar bill, that my younger sister got married on December 14, that I own a piece of property in Berkeley, and that the New York Giants won the 1991 superbowl. (Searle 1995, 1)

Institutional facts contrast with brute facts, such as „that Mount Everest has snow and ice near the summit or that hydrogen atoms have one electron, which are facts totally independent of any human opinions" (Searle 1995, 1-2).

Institutional facts differ from brute facts in their ontology: the existence of the former depends on intentionality, the latter exist independently of intentionality. And as institutional facts exist because people *believe* that they exist, *accept* that they exist, *recognize* that they exist, *expect* them to exist, *act* towards their existence, etc., it is natural that to investigate the ontology of institutional facts amounts to an investigation of those beliefs, acceptations, recognitions, expectations, actions, etc., which is to say that we have to investigate those intentional states in relation to which the institutional facts stand as their intentional objects.

In *Intentionality*, Searle emphasizes that intentional objects are just ordinary objects without there being anything special or even mysterious about their ontology. They are often the objects that exist independently of mind and the ontology of which can be described by physics, chemistry and other natural sciences. The ontology of institutional facts, however, *is* peculiar in a certain sense: they exist only as far as they are objects of intentional states. In other words: they exist only as far as they satisfy the conditions of satisfaction prescribed by the intentional states. How can they do it? How can they satisfy those conditions of satisfaction? They satisfy the conditions of satisfaction because people *think* (believe, recognize, accept, remember, etc.) that they satisfy the conditions of satisfaction. So it seems that intentionality not only prescribes the conditions of satisfaction, but is also able to "hold" the object as satisfying the conditions of satisfaction where the object itself would not satisfy them otherwise.

Searle's definition of the performance that creates and maintains institutional reality (from *The Construction of Social Reality*) goes as follows: institutional facts exist because we *collectively impose status functions to objects*. This existence-giving performance has three essential elements: the *imposition of function*, the *status* character of the imposed function, and the *collective* character of the imposition.

The imposition of function is, according to Searle, a common performance of intentionality and it is part of our everyday experience of the world: we normally experience the objects in the world as having some functions, and the functions are relative to our practical goals and interests.

> … we do not experience things *as* material objects, much less as collections of molecules. Rather, we experience a world of chairs and tables, houses and cars, lecture halls, pictures, streets, gardens, houses, and so forth. (Searle 1995, 14)

If the imposition of function is a performance of intentionality, how is it related to the accounts from the book *Intentionality*? At first sight it resembles the prescription of the conditions of satisfaction, and I think it is the prescription of the conditions of satisfaction although not in the superficial sense of saying that the object satisfies the conditions of satisfactions if it satisfies the imposed function. To prescribe the conditions of satisfaction is to determine under which aspects we experience the object. Hence we can say that the conditions of satisfaction (or more

precisely: one part of them) say: "something that I can sit on", "something that can take me from the place A to the place B", "something that I can keep water in", etc.

In the case of institutional facts, the imposed function is a *status* function. "Status function" is the title for a function that the object cannot satisfy in virtue of its natural (i.e. physical, chemical, biological, etc.) properties. While there are natural objects that can satisfy functions such as "something that can take me to the other bank of the river" regardless what (and whether) anybody accepts this satisfaction, there are no natural objects able, solely in virtue of their physical properties, to satisfy functions like "something that I can pay with" or "something that will entitle me to enter the concert hall on the concert night". The objects that in fact satisfy the functions of the latter type (i.e. money or concert tickets from our examples above) does so because a community *accepts* that they satisfy those functions. Functions of the latter type are called *status functions* by Searle, functions of the former type are called *causal functions* (or more precisely: agentive causal functions).

From the reflections above, it seems to be obvious that the intentional performance of the imposition of status function has two parts: first, the imposition of a function relative to our practical goals and interests, and second, the acceptance of an object as satisfying the imposed function. While it is just one single performance if seen from the third person point of view, there are two different intentional states from the first-personal perspective. The difference between those two states, however, is not that one of them prescribes the conditions of satisfaction and the other does something else – perhaps picks out the object directly, without any mediator: every intentional state that has a relation to reality (i.e. that has direction of fit) involves the conditions of satisfaction. What we find here are rather two different kinds of conditions of satisfaction. We have said about one of them that it prescribes the conditions of satisfaction relative to the practical goals and interests of the agent. What can we say about the other? It picks out the same object through different conditions of satisfaction, through conditions that are independent of all goals and interest of the first ones – for example through its appearance, i.e. through its shape, size, color, stiffness, etc. Then we say that you can pay with a piece of paper of such and such a shape, colors, print, size etc. Or we can pick out the object through its material, or also through its origin: you can pay with objects made of some specific material or issued in some specific way. The object can

also be identified through another function, causal or status one, and then we have an iteration of the imposition of function. All these ways are represented by the aspects of the intentional content.

Then we can say that the institutional fact is constructed in an act of identification: the thing that *has such and such properties* is identical with the thing *that allows me to do such and such*. The institutional facts have a special ontology because they are constructed as intentional objects of mental acts of some community. It seems that we can conclude that it is constructed in the performance of (at least) two intentional states of different kind: one of them is cognitive, the other one is volitive. The institutional fact is constructed as their common product.

Now we can ask: is something similar true about the intentionality that is "about" things that exist independently of it? Since the ability to impose status functions has been developed from the intentionality of a lower or more primitive level, we can suppose that the answer is "yes". Most probably, we relate to the natural objects in two different ways as well: through a function relative to our goals and interests (it must be a *causal* function in this case), and through their properties that we can perceive through our senses. Perhaps this is the essential condition under which a state can be intentional at all: to represent *the same* object in two different ways, and to do it in such a manner that it is exactly the interconnection of those two representations (i.e. the conditions of satisfaction) that presents the object as identical.

Institute of Philosophy and Religious Studies
Faculty of Arts
Charles University
Jana Palacha 2
116 38 Praha 1
Czech Republic
pavla.toracova@seznam.cz

References

SEARLE, J. R. (1983): *Intentionality. An Essay in the Philosophy of Mind.* Cambridge: Cambridge University Press.

SEARLE, J. R. (1992): *The Rediscovery of the Mind.* Cambridge (Mass.): The MIT Press.

SEARLE, J. R. (1995): *The Construction of Social Reality.* New York: The Free Press.

Searle's Defence of Internalism

Petr Koťátko

The Academy of Sciences of the Czech Republic, Prague

Abstract: The paper argues in favour of the Searlean internalist way of construing the content of thoughts and communicative acts. As interpreted by the author, it reflects the participation of external factors in the determination of content (stressed by externalists like Burge, Putnam, Kripke and Davidson) but it does so in a way which enables to approach the thinking and communicating individual as a real subject of his acts. The crucial move is to incorporate the individual's reliance on the relevant parameters of his physical or social environment into the construal of the content itself. Then the external factors can be treated as exploited by the individual himself in the articulation of his acts, rather than as intervening into the content through the gaps marked by indexicals, proper names, natural kind terms or concepts borrowed from the communal repertoire. In particular with respect to the last factor the author proposes a maxim of interpretation which he labels as The principle of maximization of subjectivity and regards as a pendant of the well-known Principle of charity.

Keywords: internalism/externalism, propositional content, indexicals, proper names, natural kind terms, division of linguistic labour.

The externalists' claim that the content of our communicative acts or of our beliefs, desires and other attitudes is – at least in some respects – determined by external factors and that these factors can play this role without any internal mediation, so that two individuals in type-identical physical and psychological states can entertain different beliefs, desires etc. and express different propositions (by uttering type-identical sentences). The externalist arguments focus on particular components of our thoughts or communicative acts which are supposed to mark gaps in the internal determination of their contents, and hence function as channels through which these external intervenes into the content take place. These components include indexicals, proper names, natu-

ral kind terms and all those general terms which are subject to the division of linguistic and intellectual labour.

John Searle suggests a complete internalist (or, as he calls it, *intentionalist* or *Fregean* – cf. e.g. Searle 1983, 22) construal of the meaning of these terms and of the propositional contents expressed by utterances of sentences including these terms. Moreover, he does so in a way which neither neglects nor obscures but rather fully reflects the ways in which our thoughts and communicative acts are embedded in their external environment, so that the externalist initiative is at least in some respects accommodated within the internalist construal of content, rather than to completely go by the board. This can be easily overlooked when we focus purely on Searle's arguments pointing to controversial presumptions behind the externalists' claims and to mistakes in their inferences.

1   An internalist reply to the externalist initiative

In the cases I am going to discuss here one can recognize a general strategy which I am inclined to call *Searlean*, although I will present it in my own words and with a motivation derived from my own priorities in the dispute with externalists. Within this strategy, the internalist addresses the externalist as follows: you claim that you have discovered external factors which, in the circumstances described by you, play an ineliminable role in the determination of content. How should we understand the way in which these factors do their job? In particular, should we suppose that the thinking or communicating subject *relies on them* in his thoughts or communicative acts? The reply which can be found in the prominent externalist theories is "yes": in fact, this is an essential part of their arguments, or at least of the way in which they attempt to make their position intuitively appealing (examples will follow soon). But then the manoeuvre which I would call the *Searlean move* follows quite naturally. If the subject relies on, or counts with, these external factors as involved in the determination of the content thought or communicated by him, oughtn't we to do justice to this reliance by reflecting it in the construal of the content itself? In particular, oughtn't we to include specification of the ways in which these external factors are supposed to do their job into the construal of the content? Obviously, this specification of the role of external factors should exploit exclusively the devices which are in the possession of the thinking subject - otherwise it could not serve to represent the subject's reliance on

these factors. But once this requirement is fulfilled, the participation of these factors in the determination of content becomes specified within the *internally fixed content* itself (within the intentional content, in John Searle's terminology) and then the contribution of the external factors does not mark any parameter or element of the content which would be left internally underdetermined. But then the externalists fail to give us a reason to postulate any content beyond (or any content broader than) internally fixed content.

In this way we can do full justice to the factors involved in the determination of content to which the externalists appeal. In general, this method of construing the content, when applied in reaction to various versions of externalism, is able to reflect the variety of ways in which our thoughts and communicative acts are anchored in their external circumstances. In fact, what could be a more consequent way of appreciating the relevance of these factors than including a specification of them and of the role they are supposed to play, into the construal of the intentional content of thoughts and communicative acts? But when doing this, we no longer present the role of these factors in the determination of content as an *intervention from the outside*: on the contrary, we present the thinker or speaker *himself* as involving these factors into the articulation of the content of his beliefs, intentions, desires etc. as well as of his assertions, promises, orders etc. In other words, we acknowledge him as a *real subject* of his thoughts and communicative acts.

To sum up, what the internalist responding in the Searlean way to the *externalist* arguments should say, when he wants to be polite, is: many thanks for a valuable innovation of the *internalist* construal of content.

2   Natural kind terms

When defending the construal of intentional content which includes a specification of the manner in which external factors are supposed to contribute to it, I have argued that this way we do justice to the thinker's or speaker's reliance on these factors. And I have pointed out that this subjective reliance is presupposed in the prominent externalist theories themselves. To begin with, let's take Putnam's theory of *natural kind terms*, e.g. his account of the term "water" as introduced into our vocabulary to refer to samples of water identified descriptively or demonstratively plus to anything else with the same essence. When jus-

tifying this account, Putnam argues that it is an essential feature of our attitude to the world that we intend our classification of natural entities to respect their essences (their internal structure responsible for the observable properties) and that we take these essences as making our applications of natural kind terms right or wrong independently on our knowledge of them (cf. e.g. Putnam 1975, 244; Putnam 1981, 46f.). But then we have a good reason to reflect this respect to essences, ascribed by Putnam to human beings, in the construal of the content of communicative acts which include utterances of the term "water" ("lemon", "tiger" etc.), as well as of the content of beliefs, desires etc. expressed in these acts. The meaning of the term "water" can then be specified e.g. by the description "anything that shares essence with what flows in rivers (on Earth), falls in drops from the clouds (on Earth) etc." – in the case that the relevant sample is picked out descriptively. Alternatively, it can be specified by the description "anything that shares essence with *this* stuff" if the relevant sample is picked out demonstratively.

The demonstrative "this", as it appears within this description, might be regarded as marking a gap within this allegedly internalist construal of content – a place through which external factors intervene into the content in a way which is not internally mediated. However, here we can appeal to Searle's internalist account of indexicals (cf. e.g. Searle 1983, 228). The resulting internalist specification of the meaning of the term "water" will then be something like: "anything that shares essence with the stuff which causes *this* experience". The demonstrative "this", as it appears within this new construction, ought not to bother the internalist at all, since it plainly refers to something internal, namely to particular experience of the subject in question (the speaker uttering the term "water").

Thus, if we want to construe the content expressed by an utterance of the sentence "Water is indispensable for life" in a way which is completely internalist (exploits only resources which are in the possession of the thinking subject) and at the same time accommodates the core of Putnam's theory, we get: "The stuff causing this experience and anything else with the same essence is indispensable for life." To sum up, Searle seems to be right when he points out in Searle (1983, 204): "Even supposing Putnam is right about his intuitions, all he has done is substitute one intentional content for another."

Even after this move one can still say, with Putnam, that meaning or intension is "extension involving" (the internal structure of particular

samples of some liquid, independently on what we know about it, enters into the principle determining what is and what is not the right application of the term "water"). However, the role of this *external* factor is fully determined in the internally fixed content of the communicative act performed in the utterance of a sentence including the term "water", and hence it is determined *internally*. It would be absurd to object: in your construal of the content you have replaced the external factor to which the externalists appeal, namely the internal structure of natural entities (e.g. the chemical composition of some stuff), by something else, namely by its description. The description "the essence of this stuff" serves precisely to identify the essence of the stuff in question, just like the description "Brown's murderer" serves to pick out a particular person (cf. Searle's polemics with Putnam in Searle 1983, 205f.).

3 Proper names

Similarly, *Kripke's theory of proper names* not only links our utterances of proper names to individuals as their referents through the chain of uses of the name, anchored in the initial baptism (cf. e.g. Kripke 1972, 91f.). It appeals to the *intention* of any user of the name to join the chain – more specifically, to refer, when uttering the name, to the same individual as his predecessor in the chain (cf. Kripke 1972, 96f., 162). This is an essential part of the theory, since this is supposed to connect one segment of the chain with another, and so to guarantee the continuity of the chain and its ability to link the name to its referent. What could then be more natural than to reflect this speaker's respect to the (Kripkean) referential mechanism in the construal of the content expressed by his utterance of a sentence including a name? The intentional content supposed to do the identificatory work in a case in which the speaker utters the name 'Jan Novák' with the plain intention to refer to the same person as the speaker from whom he got the name, without having any independent identifying knowledge, can be specified precisely by means of the description "the man referred to by the speaker from whom I got the name 'Jan Novák'".[1]

_____

[1]   If the speaker possesses some non-parasitic identifying knowledge, it is natural to suppose that it has the form of a *cluster* of identifying descriptions – details see in Searle (1958); Searle (1969, Ch. 7); cf. also Zsofia Zvolenszky's paper in this volume.

This is simply a way of reflecting (one of the possible forms of) the parasitic uses of names (cf. e.g. Strawson 1959, Ch. 6; Searle 1983, Ch. 9). But surely we can in the same (descriptivist) way adopt the whole Kripkean picture of the referential role of names. The intentional content supposed to pick out the referent of a particular utterance of the name 'Jan Novák' (by a speaker not equipped with any independent, non-parasitic identifying knowledge) can be specified e.g. by means of the description: "the individual baptized by the name 'Jan Novák' at the beginning of the chain to which this utterance belongs".[2] The reason why the description should include not only reference to the name but also to its particular utterance, is that it serves to identify the relevant chain among (supposedly) thousands of chains in which the name 'Jan Novák' appears.

To recall Brian Loar's aphorism (from Loar 1980) , this is a way in which the causal theory of names acquires "self-consciousness". The functioning of the mechanism which links names with their referents is presented as essentially including deliberate reliance of language users on that very mechanism. This move does not weaken or blur the *external* reality of the social mechanism in question: however, from this new point of view, the speaker *himself* introduces the mechanism into the articulation of the content of his communicative act (and of the thought expressed in it). In other words, the mechanism does not apply from the outside on the speaker's act; rather, the act includes an *exploitation* of the mechanism in its articulation.

4   The social conceptual repertoire

The same kind of move, with the same motivation, should be applied to the externalist account of the role of social conceptual repertoire in the articulation of our thought contents. Let me focus on Tylor Burge's famous arthritis example, which is not discussed in Searle (1983) , unlike Burge's early externalist theory of *de re* beliefs (beliefs characterized by Burge as including ineliminable indexical elements; cf. Burge 1977), Putnam's *Twin-Earth* thought experiment and Kripke's causal theory of proper names. Nevertheless it should be, as I have just suggested, approached in the same way.

_____

2   Kripke himself admits such a possibility (and attempts to diminish its relevance) in a brief footnote in Kripke (1972, 88, cf. also 162).

In fact, we can, for our purposes, get along with the situation described in the first part of Burge's notoriously known example (from Burge 1979). There we learn that somebody called Bert has not fully mastered the standard notion of arthritis as it is used in our medical science. Our evidence for this is that he seriously and sincerely utters sentences like "I have arthritis in my thigh". According to Burge, this fact in itself should not prevent us from using the term "arthritis" quite straightforwardly in attributing beliefs and other attitudes to Bert. In particular, we have full right to apply disquotation on the utterance mentioned, with the result: "Bert believes that he has arthritis in his thigh." When doing so, we (provided that we use that sentence in the standard way) project the conventional meaning of the term "arthritis", i.e. the medical concept of arthritis, into the content of Bert's belief - although we know precisely about this concept that it has not been mastered by Bert.

This is quite a radical move, since in sentences like "Bert believes that he has arthritis in his thigh" the term "arthritis" occurs in the so called *oblique* position, which means that its replacement by another term identifying the same disease under another mode of presentation can change the truth value of the whole attribution. That shows that in such a sentence we are identifying the content of Bert's belief not just by relating it to certain disease but by specifying the mode of presentation under which he is thinking about it.[3] And that means that the communal conceptual repertoire fixed by the conventions of our sociolect is allowed to intervene into the apparently most internal sphere of thought content, to erase there what is incompatible with the standard and to supply what is missing. The principle seems to be that even those aspects of the concepts fixed in a given community, which have not been mastered by the subject, can participate in the articulation of his attitudes - in virtue of his having the status of a competent member of that community.

When arguing for this position, Burge (1979, 101, 114) points to an important empirical phenomenon, made popular by Putnam under the title "division of linguistic labour" (cf. e.g. Putnam 1975): individuals typically do respect the authority of experts, concerning meanings of

_____

[3] Cf. Burge (1979, 114): "Clearly oblique occurences in mentalistic discourse have something to do with characterzing a person's epistemic perspective – how things seem to him or how they are represented to him."

words and conceptual components of the contents of their attitudes. In these issues, like in many others, individuals feel obliged to the social standards even if they have not fully mastered them. To put it positively, even those concepts from communal repertoire which we have not mastered can be engaged in the articulation of our attitudes, and hence in the determination of their satisfaction conditions. This is the profit we gain from the division of linguistic and intellectual labour, if we participate in it and rely on it.

The question is whether the submissivity to experts or to social standards provides us with any argument in favour of social externalism.[4] As I have pointed out, John Searle does not discuss Burge's arthritis example in Searle (1983), even though he does comment there on the idea of the division of linguistic labour. Yet the Searlean manner of facing externalist arguments is well applicable here. If we admit that Burge's Bert, in the articulation of his belief, relies on (or defers to) the medical definition of arthritis unknown to him, the internalist should insist that this reliance should be explicitly reflected in our specification of the content of Bert's belief. The simplest and most natural way of doing so is to say: "Bert believes that he has in his thigh the disease referred to by experts in his community as 'arthritis'." Or: "Bert believes that he has in his thigh a disease satisfying the medical notion of arthritis." Or you can opt for some other formulation with the same effect – namely of directly involving Bert's deference to experts into the construal of the content of his belief.

5   Some problems

Obviously, the value of such constructions largely depends on your ability to explain their philosophical implications, to demonstrate their applicability on cases which may be found controversial, and to defend them against objections of all possible kinds. It is here that the detailed analytical work starts. Let me just mention some important issues without going into them:[5]

---

[4]   An opposite view has been expressed e.g. in Searle (1983, 201f), Bilgrami (1992, 67f), Davidson (1994, 5).

[5]   I have attempted to do so in Koťátko (2006a) and in more detail in Koťátko (2006b, Chap. C.III.2).

a) The analysis must be demonstrably compatible with the assumption that the deferring layperson to whom it is meant to apply can share beliefs about arthritis with an expert, despite the fact that the content of the latter's belief does not include any deference to the expert use of the term "arthritis" or to the expert notion of arthritis.

b) The same concerns the assumption that English and Czech laypersons can share beliefs about arthritis, even though the metalinguistic part of their belief contents includes, according to the analysis, reference to two different words: "arthritis" and "artritida".[6]

c) It should be made clear that the metalinguistic analysis does not turn beliefs about non-linguistic entities into beliefs about language (as T. Burge objects in 1979, 97).

d) The analysis must not imply that the sentences "I have arthritis in my thigh" and "I have in my thigh the disease referred to by experts in this community as 'arthritis'" have the same meaning.

e) The metalinguistic analysis must not lead to the consequence that statements like "Arthritis is called 'arthritis' by experts in this community" come out as necessarily true: rather, the analysis should allow us to treat them as examples of the *contingent a priori*.

f) It should be made clear that the analysis does not generate any destructive kind of regress; in particular, that it is not the case that once it is applied at a certain level its completeness requires that it also be applied at a higher level, i.e. to the expressions employed in it, etc. *ad infinitum*.[7]

g) The analysis should account for all possible distributions of the roles of the layperson and of the expert between the interpreted person and the interpreter (and of their attributions of these roles to one another). For different combinations, the structure of the belief attribution (and within it the construal of the attributed belief) will also differ.[8]

---

[6] T. Burge finds this problem quite serious (cf. Burge 1979, 96), G. Segal points to it in one of his arguments against the metalinguistic construal of belief contents (see Segal 2005, 115), H. Putnam has done the same in the discussion at the Karlovy Vary Symposium on *Swimming in XYZ* in 1998.

[7] Gabriel Segal has argued that it is the case in Segal (2000, 115). I have offered arguments against this view in Koťátko (2006, 63f.).

[8] I have attempted to say more about this in Koťátko (2010, Chap. 3).

h) The analysis should take into account the possibility of "switches" in which the interpreted person moves from one social environment to another, belonging to the same linguistic community but varying in the conceptual repertoire and in the standards governing the use of some terms.[9]

6 Some implications

The defence and elaboration of the internalist construction of the thought content with respect to these and similar issues is perhaps the most attractive part of the project – but here we have to remain on the level of the basic philosophical motivation and implications of the internalist position. In general, the kind of the construal of the thought content defended in Chapter 4 can be viewed as an internalist way of appreciating the social parameters of thought to which philosophers like Burge or Putnam have drawn our attention. If we admit that linguistic conventions or the expert usage participate (at least in some cases) in the determination of the content of our beliefs and other attitudes as radically as Burge insists, we should have no problem with the idea that in such cases the belief content itself includes an appeal to these conventions or to the expert usage. The function of this appeal is to bridge the gap between the expert meanings or communal conceptual repertoire and the actual linguistic or intellectual equipment of the individual in question. In other words, it enables to the individual subject to exploit (in the articulation of his beliefs) even those components or aspects of the standard communal conceptual repertoire which he has not mastered. To reject this while accepting the socialist claims about the nature of our language and thought (in Burge's sense) amounts to insisting that individuals are rather *objects* than *subjects* of the division of linguistic and intellectual labour; or that the integration of individuals into the community on this level is rather *external unification* than *active participation*; or that what is going on here is some kind of *self-projection* of communal concepts into the content of the subject's beliefs rather than the subject's *exploitation* of the concepts from the communal repertoire.

---

[9] On the phenomenon of switches and its implications cf. e.g. Warfield (1987), Ludlow (1985), Ludlow (1987).

From the internalist point of view, in the Bert-like cases we have just two options. Either we feel justified to ascribe deference to social standards to the subject in question, and then we should include it in the construction of his beliefs; or we see no reasons to assume deference, and then we are obliged to do our best to identify the subject's notions, including their idiosyncratic features, on the basis of his linguistic as well as non-linguistic behaviour. The third option, namely to adopt the position of managers (or wardens) of the sociolect and of the communal conceptual repertoire fixed in it and to approach the subject's thought contents as liable to our tutelage, is simply not open to us.

The externalism-internalism dispute concerning the social nature of thought provides a good opportunity to ask again the old question what it is for an individual to *have* a concept or to *have* a belief with a certain content. From the point of view of social externalists, having a concept does not necessarily mean knowing the criteria determining its extension (and knowing the application rules of the corresponding term). That implies that an individual can have a belief with quite determinate content without being able to specify its truth-conditions: it is enough that these conditions are determined on the level of the communal conceptual repertoire. From the viewpoint of the socially sensitive internalist it is not so: we can keep the intuitive assumption that to have a concept is to know the criteria of its application (in sense of a complete set of necessary and jointly sufficient conditions), but these criteria can be *parasitic* or *deferential*. If the layperson's notion of arthritis has the form "the disease referred to by experts as 'arthritis'", it obviously gives a criterion unambiguously determining the extension, provided that the appeal to experts is satisfied, i.e. provided that in the given community there is precisely one expert meaning of the term 'arthritis' which unambiguously determines its extension. Granted this, the layperson's belief which would be properly expressed by the sentence "Arthritis is infectious" has unambiguous and internally fixed truth conditions.

From this perspective, having a belief and having a concept seems to be a more respectable role than it appears to be from the point of view of social externalists. But not necessarily much more. The deference to experts or other kinds of authorities gives them in any case some power over the content of our beliefs. For instance, a man reading in his favourite newspaper the sentence "Extremists (internalists, individualists etc.) should be made harmless" may find it appealing and

adopt a belief construed as follows: "Those who are classified by competent authorities as 'extremists' should be treated in a way referred to by competent authorities as 'making somebody harmless'." In such cases the thinking subject "retreats" from the content of his own attitude or communicative act and delegates its articulation on concepts which are in possession of others. Even this delegation can be, if we are charitable enough, interpreted as a way of the subject's involvement in (or even his control over) the articulation of the content of what he thinks or communicates – but this can hardly impress an individualist in the theory of content. The most natural thing to say about the individual from our last example is that he allows his own subjectivity to dissolve in the external collective mind (controlled by the authorities). In an extreme case, the whole conceptual work of such an individual can consist in the identification of (and appeal to) external resources of the articulation of his own thoughts and communicative acts – and then we have full right to say that the thoughts of such an individual "pass him by".

According to the individualist position I am defending here, the interpretation should treat the interpreted subject as finding himself in such an embarrassing position only if there is no other option compatible with the available evidence. In other words, the interpretation should be allowed to weaken the subject's control over the content of his own thoughts and communicative acts only under irresistible pressure of evidence. Or to put it positively: the idea of a subject maximally involved in the articulation of his attitudes and communicative acts should play the same regulative role as the idea of a coherently and rationally thinking and acting individual. To say so is to commit oneself to a certain version of the well-known *principle of charity*. I will call it the *principle of the maximization of subjectivity*, and (inspired by Kant's presentation of his categorical imperative) submit it in three equivalent formulations:

a) Do not ascribe to the interpreted person P contents (of thoughts or communicative acts) such that their articulation depends on conceptual resources beyond the capacities of P, whenever the available evidence (the communicative as well as the non-communicative behaviour of P) admits an alternative interpretation.

b) Among the alternative ways – all equally supported by the available evidence – of identifying the contents thought or communicated by the person P, choose the one which maximizes P's subjective involvement in his attitudes or acts.

c) The interpreted person should not, as a result of the interpretation, come out as more dependent on the division of linguistic and intellectual labour than is needed for making sense of his overall behaviour.

Institute of Philosophy
The Academy of Sciences of the Czech Republic, *v. v. i.*
Jilská 1
110 00 Praha 1
Czech Republic
kotatko@flu.cas.cz

References

BILGRAMI, A. (1992): *Belief and Meaning*. Oxford: Blackwell.

BURGE, T. (1977): Belief De Re. *Journal of Philosophy* 74, pp.

BURGE, T. (1979): Individualism and the Mental. In: French, P. – Uehling, T. – Wettstein, H. (eds.): *Midwest Studies in Philosophy, IV: Studies in Metaphysics*. Minneapolis: University of Minnesota Press.

DAVIDSON, D. (1994): The Social Aspect of Language. In: McGuiness, B. – Oliveri, G. (eds.): *The Philosophy of Michael Dummett*. Dordrecht – Boston – London: Kluwer Academic Publishers.

KOŤÁTKO, P. (2006): Having a Concept. In: Marvan, T. (ed.): *What Determines Content? The Internalism/Externalism Dispute*. Newcastle: Cambridge Scholar Press, 55-68.

KOŤÁTKO, P. (2010): Realismus a jeho implikace. [Realism and Its Implications] In: Zouhar, M. (ed.): *Realismus, internalizmus, individualismus. Ku knihe Petera Koťátka Interpretace a subjektivita*. Bratislava: Infopress, 11-30.

KRIPKE, S. (1972): *Naming and Necessity*. Oxford: Blackwell

LOAR, B. (1980): Names and Descriptions. A Reply to Michael Devitt. *Philosophical Studies*, 38, 85-89.

LUDLOW, P. (1985): Externalism, Self-Knowledge and the Prevalence of Slow Switching. *Analysis* 55, pp.

LUDLOW, P. (1987): On the Relevance of Slow Switching. *Analysis* 57, pp.

PUTNAM, H. (1975): The Meaning of „Meaning". In: Putnam, H.: *Mind, Language and Reality*. Cambridge: Cambridge University Press.

PUTNAM, H. (1981): *Reason, Truth and History. Cambridge*: Cambridge University Press.

SEARLE, J. R. (1958): Proper Names. *Mind* 67, pp.

SEARLE, J. R. (1969): *Speech Acts. An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

SEARLE, J. R. (1983): *Intentionality. En Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.

SEARLE, J. R. (1995): *The Construction of Social Reality*. London: Penguin.

SEGAL, G. (2000): *A Slim Book about Narrow Content*. Cambridge (Mass.): The MIT Press.

STRAWSON, P. F. (1959): *Individuals. An Essay in Descriptive Metaphysics*. London: Routledge.

WARFIELD, A.T. (1987): Externalism, Priviledged Self-Knowledge and the Irrelevance of Slow Switching. *Analysis* 57, pp.

III. Philosophy of language

Searle on Analyticity, Necessity, and Proper Names

Zsófia Zvolenszky

Eötvös University, Budapest

Abstract: My aim is to show that once we appreciate how Searle (1958) fills in the details of his account of proper names – which I will dub the presuppositional view – and how we might supplement it further, we are in for a twofold discovery. First, Searle's account is crucially unlike the so-called cluster-of-descriptions view, which many philosophers take Searle to have held. Second, the presuppositional view he did hold is interesting, plausible, and worthy of serious reconsideration. The idea that Searle's account is a largely Fregean interlude between the Fregean description theory of proper names and Kripke's proposals presented in "Naming and Necessity" is in major ways a myth, a mythical chapter in how the story of 20th-century philosophy of language is often told.

Keywords: proper names, meaning, reference, definite descriptions, presuppositions.

1   Introduction

John Searle begins his seminal paper "Proper Names" with the question: "Do proper names have senses?" (Searle 1958, 166). He sums up part of his reply towards the end of the paper: if the question "asks whether or not proper names are logically connected with characteristics of the object to which they refer the answer is 'yes, in a loose sort of way'" (Searle 1958, 173). He briefly mentions at this point that the logical connections involve "descriptive presuppositions" (cf. Searle 1958, 173): various descriptions that capture characteristics of the object, uniquely identifying it; for example, the descriptive presuppositions for the name 'Aristotle' might include 'the teacher of Alexander the

Great', 'the most famous pupil of Plato', 'the author of The Metaphys-
ics'. These descriptions are supposed to express certain characteristics
of Aristotle: his having been the one and only teacher of Alexander
the Great, for example.[1] That such descriptions are featured in, of all
things, *presuppositions*, is a widely ignored feature of Searle's proposal,
which I will argue is nonetheless crucial and innovative; indeed, it's
so central that I will refer to Searle's proposal as the presuppositional
view.

My aim is to show that once we appreciate how Searle fills in the
details of his presuppositional view and how we might supplement it
further, we are in for a twofold discovery: first, Searle's (1958) account
is crucially unlike the so-called cluster-of-descriptions view, which
many philosophers take Searle to have held; and second, the presup-
positional account he did hold is interesting, plausible, and worthy
of serious reconsideration. The idea that Searle's account is a largely
Fregean interlude between the Fregean description theory of proper
names and Kripke's proposals presented in "Naming and Necessity"
is in major ways a myth, a mythical chapter in how the story of 20th-
century philosophy of language is often told. Contrary to philosophical
lore, a Searlean theory bears close kinship to direct reference theory, a
view inspired by Kripke, with descriptive presuppositions adding an
interesting twist.

In the course of this paper, I will revisit Searle's (1958) "Proper
Names" to expose the presuppositional view that he proposed there. I'd
like to show that this view can be developed further and can be defended
against some central objections raised against the cluster-of-descriptions
view of proper names. After providing some background (Section 2),
I will argue that Searle's view was misunderstood in part because it wasn't
recognized that by 'necessity', Searle meant 'analyticity'(Section 3),
and in part because the role of presuppositions has been ignored (Sec-
tion 4). These considerations already offer responses to some of the
objections raised against the cluster-of-descriptions view. Meanwhile
a Searlean framework also shows considerable promise in accounting
for some of the remaining objections, concerning singular existential
claims like 'Aristotle existed' (Section 5). Concluding remarks will fol-
low in Section 6.

---

[1] Of course, some characteristics of Aristotle's are not unique to him, like
being a philosopher, and accordingly, the description expressing it, 'a phi-
losopher' is not a *definite* description.

2 Three commonly raised objections

Frege's (1892/1952) reply to our initial question had been: yes, proper names do indeed have senses, what we might call Fregean meanings. *Fregean meanings* are usually characterized as playing a number of theoretical roles including these: specifying what proper names contribute to the meanings of complex expressions containing them; determining to whom or (to what) the name refers; and being the objects of understanding (Fregean meanings are what competent users of the name grasp). The dominant view among philosophers is that Frege's theory of name meaning – the description theory of the meaning of proper names – provides an elegant solution to these (and other) tasks, but encounters fatal problems that a patched-up amendment, the cluster-of-descriptions theory fails to resolve. It is this latter theory, attributed primarily to Searle, that is therefore commonly featured in the literature as a lead-up to and foil for Saul Kripke's alternative proposals about how proper names work. In this section, I will outline the Fregean theory and its cluster-based successor, thought to be Searle's, and how they are taken to fare with respect to three objections.

The description theory of the meaning of proper names (the *description view* for short) – attributed to Frege (and also to Russell) – holds that for each proper name, its meaning is given by an associated definite description. For example, the meaning of 'Aristotle' is given by an associated description like 'the teacher of Alexander the Great'. Who does the associating – the linguistic community, or individual speakers? The Fregean theory claims the latter: each and every speaker who is a competent user of 'Aristotle' associates with the name the description 'the teacher of Alexander the Great'. And what does the associating consist in? There are various ways we might go on this, a convenient choice is to follow Kripke's (1970/1980, 64, 71) characterization: a speaker associates a definite description with a proper name just in case she believes that the description fits a single individual – the bearer of the name.[2] According to the description view, understanding the name 'Aristotle' requires associating 'the teacher of Alexander the Great' with it; it is the meaning of this description that the name contributes to sentences containing it; and the bearer of the name is whoever fits the

---

[2] This is featured in theses (1) and (2) of Kripke's general characterization of description-based accounts of which the description view discussed here is a special case.

definite description. The name, in short, is synonymous with the associated definite description.

The objections I am about to discuss are commonly raised against the description view. In what follows, I am closely tracking Michael Devitt and Kim Sterelny's (1999, 45-54) as well as Peter Ludlow's (2007, Section 8) line of argument; both of these texts are highly influential reference points and teaching materials within the philosophy of language.

People typically associate a plurality of definite descriptions with a name. There is, on the one hand, intrapersonal plurality: a philosopher might associate with 'Aristotle' the descriptions 'the most famous student of Plato', 'the author of The Metaphysics', 'the author of The Nichomachean Ethics'. And there is, on the other hand, interpersonal plurality: a historian might associate with 'Aristotle' 'the teacher of Alexander the Great', a description that is different from the philosopher's or from a high school student's.

(i)   *The principled basis objection* concerns both intrapersonal and interpersonal plurality. The description theory posits a single definite description as giving the meaning or sense of 'Aristotle'. But for a single speaker and across several speakers, we find a plurality of descriptions associated with the name. According to the description view, "…one of these descriptions trumps all the others. If it fails to denote, the name is empty, even if all the other associated descriptions pick out the one object" (Devitt – Sterelny 1999, 48). It is up to the description view to provide a principled basis for selecting *the* definite description – the *meaning-giving description* – with which the name 'Aristotle' is to be synonymous; but it's unclear whence that principled basis would come.

(ii)   *The unwanted ambiguity objection* concerns interpersonal plurality. Even if we were able to respond to the principled basis objection with respect to intrapersonal plurality, the meaning-giving descriptions will most likely vary across speakers within a linguistic community. The philosopher's, the historian's, the high school student's, etc. meaning-giving description associated with 'Aristotle' are unlikely to be the same. But then the description theory has it that 'Aristotle' is multiply ambiguous among language users, an unwelcome result.

(iii) *The unwanted necessity objection* makes the point that even if objections (i) and (ii) were handled, the description theory can't be squared with the fact that proper names and the candidates for meaning-giving descriptions we have thus far considered have distinct modal profiles.

(1)   Aristotle taught someone.
(2)   The teacher of Alexander the Great taught someone.

(1) is a contingent truth while (2) is necessary. It is a contingent fact about Aristotle that he taught someone; in some counterfactual circumstances, he didn't teach anyone at all. By contrast, in all counterfactual circumstances, whoever fits the description 'the teacher of Alexander' will be a person (creature) who taught someone, making (2) necessary.[3] But positing 'the teacher of Alexander' as the meaning-giving description for 'Aristotle' would wrongly predict (1) to be a necessary truth. And any of the candidates for meaning-giving descriptions considered so far would give rise to similar unwanted necessary truths.

The following amendment of the description view seems at first to come to the rescue: the meaning of a proper name like 'Aristotle' is given not by a single definite description, but instead by a cluster of descriptions. Call this view, incorporating the following four features, the *cluster view*, which philosophers have widely attributed to Searle. First, the cluster for 'Aristotle', say, might include elements that don't pick out Aristotle uniquely: philosopher, was born in Stagira. Second, the cluster is gleaned from across speakers: from the philosopher, the historian, the high school student, and so on. Third, it is enough that individual speakers associate with the name 'Aristotle' some or other description or combination of descriptions from the cluster such that for each speaker, the majority (or weighted majority) of her descriptions uniquely identify Aristotle; all such speakers then count as using and understanding one and the same name 'Aristotle'.[4] Fourth, for a

---

[3]   Of course, we are talking about the so-called *de dicto* reading of (2) in which the definite description takes narrower scope than the modal auxiliary 'might'. The issue is: proper names and definite descriptions have distinct modal profiles in that the latter produce a *de dicto* reading while the former do not. For thorough discussion of the ways in which a proponent of the description view can appeal to the two readings and what problems she encounters in the process, see Soames (2002, 24-50).

[4]   Devitt and Sterelny (1997, 50) also mention a weaker variant of this third feature: to use and understand the name 'Aristotle', it is sufficient that a speaker associate *some* definite description the (weighted) majority of which uniquely identifies the object; it need not be a description in the cluster, it could instead be something like 'whoever John was referring to just now with his use of "Aristotle"'. On this weaker variant, the description cluster gives the meaning of the name at the level of the linguistic community as

person to be the bearer of the name 'Aristotle', it is enough that most (or a weighted most) of the descriptions in the cluster fit him; the cluster view can accommodate some degree of error within the cluster (say, if we find out that Aristotle wasn't born in Stagira after all), as long as the (weighted) majority of the descriptions fit a certain individual, he counts as the bearer of the name.

Initially, the cluster view seems to have responses ready to objections (i)-(iii). In response to (i): The cluster includes several descriptions, so no principled basis is needed to choose just one among them. In response to (ii): no unwanted ambiguity arises for 'Aristotle', because the historian's, philosopher's and others' associated descriptions are all included in the cluster for the name. And in response to (iii): we avoid unwanted necessities like (1) because the cluster view does not require that *all* descriptions in the cluster fit the bearer of the name: in

───────────────────────────────

a whole, but doesn't play a role in speaker understanding. Plausibly, Strawson (1959, 181-183) held such a variant of the cluster view. The objections and arguments considered in this paper apply equally to both formulations of this third feature.

It is, however, worth mentioning briefly how some of Kripke's objections to description-based views apply to something like Strawson's view with the weak formulation of the third feature. Crucially, Strawson posits the cluster of descriptions at the level of the community. Meanwhile, according to him, an individual's uniquely identifying description (which need not be included in the cluster) is, clearly, at the level of the individual. We can thus interpret Strawson as having given not one but two description-based proposals: *a community-level account about proper name meaning (in terms of description clusters), and an individual-level account of proper name use and understanding*. (By contrast, the Fregean description view is a unified individual-level account of meaning, use and understanding.) In Lecture II, Kripke (1970/1980) raises various problems that he thinks are applicable to all description-based theories of proper names: a competent user of a name need not associate any definite description with the name; and even if she does, her description might be in error and fit no-one or fit someone other than the bearer of the name. Now, these objections concern only individual-level description-based accounts; they leave untouched community-level alternatives. This way, one of Strawson's proposals—about community-level clusters—is unaffected by the just-mentioned Kripkean objections (as Evans 1973, 187-189 points out), yet (*pace* Evans) the Strawsonian is faced with these Kripkean objections when it comes to his other proposal about uniquely identifying descriptions at the level of the individual.

As we'll see shortly, the principled basis objection is distinctive in that it applies even to community-level versions of description-based views.

a counterfactual circumstance, 'the teacher of Alexander the Great' is one of the minority descriptions that fails to fit the bearer of the name, making (1), "Aristotle taught someone", a contingent truth, just like we wanted.

But on closer inspection, objections (i)-(iii) return to haunt the cluster view as well.

The principled basis objection remains: we don't want every single associated description about Aristotle to go into the cluster: line-by-line details about his writings that an Aristotle-scholar might cite plausibly aren't part of the cluster that *defines*, gives the meaning of the name. Nor do we want to include in the cluster someone's idiosyncratic description like 'the philosopher I kept calling "Aristid" in my philosophy final exam'. Peter F. Strawson, one of the philosophers to whom the cluster view is attributed, suggests that we cull the cluster of descriptions by asking individual speakers for what they consider to be "salient" descriptions about Aristotle, incorporating in the cluster "the most frequently mentioned" ones (Strawson 1959, 191). The cluster view therefore still has to provide a principled basis for separating what's salient and frequently mentioned from what isn't.[5]

The unwanted ambiguity objection returns: it is unlikely that the clusters of descriptions and the relative weight assignments should be the same across individuals, the historian, the high school student, etc. But that still yields the unwelcome result that 'Aristotle' is ambiguous among these speakers, has different meanings across language users. If the proponent of the cluster view tried to make amends by suggesting that the linguistic community fixes the cluster across speakers, then one of the prime advantages of the description view and the cluster view would vanish: their ability to explain the contrast between informative identity statements like 'Cicero is Tully' and trivial ones like 'Tully is Tully' (the associated descriptions for 'Cicero' and 'Tully' are distinct, resulting in a difference in meaning—the explanation goes). For it is likely that the cluster of descriptions and relative weights culled from the linguistic community as a whole would be identical for the two

---

[5]   Notice that here, the principled basis objection targets Strawson's community-level proposal of description clusters: on what basis do we cull the cluster from across speakers? See the previous footnote on community- versus individual-level accounts.

proper names 'Cicero' and 'Tully', making the first identity statement trivial like the second one.

The cluster view still has to contend with unwanted necessities:

(3)   Aristotle did at least one of the deeds featured in the cluster.

(3) is plausibly contingent: Aristotle's life might have gone entirely differently, he might have chosen a different profession, so none of the deeds mentioned in the cluster are true of him. Yet the cluster view wrongly predicts (3) to be a necessary truth.

Arguing along these lines, many authors – including Devitt and Sterelny as well as Ludlow – use the cluster view as an interlude to lead up to Kripke's 1970 "Naming and Necessity" lectures (see Kripke 1970/1980) which introduced alternatives to the description and the cluster views. But is the cluster view the one Searle (1958) put forth? And is Searle's own proposal subject to objections (i)-(iii), as philosophers have tended to assume? In what follows, I will motivate a negative answer to both these questions.

3   Necessity and analyticity

In this section, I argue that Searle's (1958) theory in "Proper Names" does avoid (iii), the unwanted necessity objection. Philosophers have thought otherwise because they understood necessity differently than Searle did.

As a starting point, it is well to clarify what Searle means by 'analytic', a notion with which he begins and ends his paper. According to him, "[a] statement is analytic just in case it is true in virtue of linguistic rules alone, without any recourse to empirical investigation" (Searle 1958, 166). Now, given this, do we have reason to think that (4) is synthetic?

(4)   Tully is Cicero.

Searle thinks not, on the grounds that (4) *can be used* in such a way that it "follow[s] from linguistic rules"; later on he adds: "[a] statement made using this sentence would … be analytic for most people" (Searle 1958, 167, 173). When used analytically, (4) provides information about how linguistic rules governing the use of various symbols of English like 'Cicero' and 'Tully' are: these rules make it so the two names pick out the same individual. Meanwhile, (4) can be used to make a synthet-

ic statement also; "it might even advance a historical discovery of the first importance" (Searle 1958, 173). Here, Searle likens (4) to debates over the truth of 'Shakespeare is Bacon', which he takes to be a synthetic statement: participants in this debate "are not advancing a thesis about language" (Searle 1958, 167), and hence aren't making an analytic statement when saying or denying 'Shakespeare is Bacon'. In "Proper Names", Searle sets out "to examine the connection between proper names and their referents in such a manner as to show how both kinds of identity statement [analytic and synthetic] are possible" (Searle 1958, 167). In effect, Searle is outlining a notion of analytic-for-an-individual or analytic-for-a-subgroup rather than analytic-for-an-entire-linguistic-community. Notice that in the preceding quotes, he had talked about 'Shakespeare is Bacon' being used by *debate participants* to make a synthetic statement, and (4) being "analytic for most people", but synthetic for some.[6] It is unclear what purpose this individual-relativized (or subgroup-relativized) notion would serve; I will return to this issue briefly in Section 5. In the rest of this section, I'll explore how Searle's notion of necessity relates to his notion of analyticity.

The oft-quoted passage[7] from "Proper Names" in connection with the unwanted necessity objection goes as follows:

> … suppose we ask, 'why do we have proper names at all?' Obviously, to refer to individuals. 'Yes, but descriptions could do that for us.' But only at the cost of specifying identity conditions every time reference is made: suppose we agree to drop 'Aristotle' and use, say 'the teacher of Alexander', then it is a **necessary truth** that the man referred to is Alexander's teacher—but it is a **contingent fact** that Aristotle ever went into pedagogy (though I am suggesting it is a **necessary fact** that Aristotle has the logical sum, inclusive disjunction, of properties commonly attributed to him: any individual

---

6   Searle doesn't mention the possibility of a single individual using (4) at one time to make an analytic statement, and at another, to make a synthetic one. Based on his remarks about analyticity, it seems clear that he would definitely allow such a possibility if the linguistic rules for 'Cicero' and 'Tully' were to change over time. What about other scenarios in which the linguistic rules remain constant? This turns out to be an interesting issue that I will address in Section 5.

7   Kripke quotes the bulk of this passage twice (1970/1980, 61, 74) while Devitt – Sterelny quote the parenthetical remark (1999, 51).

> not having at least some of these properties could not be Aristotle).
> (Searle 1958, 172; underlining, boldfacing added)

By 'contingent'/'necessary', occurring here, Searle plausibly means a linguistic notion: '*contingent/necessary given linguistic rules*'; there is no reason to interpret him in any other way. In the parenthetical remark, he suggests that *Aristotle in the actual world* is bound to have at least one of the properties commonly attributed to him. Searle isn't talking about *Aristotle in a counterfactual circumstance* in which he becomes a carpenter, say. The unwanted necessity objection hinges on the latter interpretation of 'necessary fact', and hence doesn't arise against Searle. In sum, by 'necessity', Searle – like most of his contemporaries[8] – plausibly meant 'analyticity'. As telling evidence for this interpretation, consider the fact that this passage is repeated almost verbatim in Searle's book *Speech Acts* a decade later, with 'necessary truth' replaced by 'analytic truth' (cf. Searle 1969, 172-173).

The unwanted necessity objection involves a different notion of necessity, that of metaphysical necessity, clarified by Kripke in his "Naming and Necessity" lectures in which he returns to the above passage from Searle twice, suggesting that *if* in the passage 'necessary' means metaphysical necessity, *then* Searle's parenthetical remark is false, for it isn't "a necessary truth that Aristotle had the properties commonly attributed to him". This is a conditional form of the unwanted necessity objection then. Kripke and subsequent commentators like Devitt – Sterelny (1999) as well as Ludlow (2007) took the conditional antecedent to be true, reading Kripke's notion of metaphysical necessity into Searle's writing, a move I hope to have shown is unfounded.

Elsewhere in "Proper Names", Searle does seem to use 'contingent' (three times) in a different sense. But I aim to show that this only serves to reinforce and not weaken the point I've been making: that by 'contingent/necessary', Searle means 'contingent/necessary given linguistic rules'. Towards the very beginning of the paper, Searle makes the point that (4) can be used to make an analytic statement that nonetheless carries information, to wit, information about the linguistic rules for the symbols (for example, the proper names) of our language. Sear-

---

[8]   See for example the debate between Marcus (1961) and Quine (1963) as well as some of Quine's earlier work (e.g. 1943, 1953). For an excellent overview of debates over and changes in the notion of the necessity up until the 1970s, see Burgess (1997) and Neale (2000).

le emphasizes that when (4) is so used, it is nonetheless contingent, "illustrat[ing] contingent facts about our use of symbols", (4)'s truth is a matter of "contingent usage" (Searle 1958, 166, 167). There are three things worth noting about this notion of contingency-given-how-usage-is (*usage-contingency* for short). First, usage-contingency is closely related to Kripke's metaphysical notion of contingency—about counterfactual circumstances in which linguistic rules develop differently, say, 'is' means what 'loves' actually means.[9] Second, in the usage-contingency sense, every use of every sentence is contingent, for the linguistic rules for all expressions are a matter of how language use happens to have developed; it's overwhelmingly plausible to expect that no statements or facts are ever usage-necessary. And given this, third, it is clear that in the widely quoted passage from Searle, above, 'contingent fact' and 'necessary fact' are used in a sense different from the usage-contingency sense: after all, Searle writes that "it is a necessary fact that Aristotle has the logical sum, inclusive disjunction, of properties commonly attributed to him", yet as we have just noted, there are no *usage*-necessary facts; it is a *usage-contingent* fact that certain properties (and not others) are attributed to the bearer of 'Aristotle'. In sum, although Searle employs a different notion of contingency elsewhere in his paper—what I have dubbed usage-contingency—upon closer inspection, it is obvi-

---

[9]   Two remarks are in order, one about Searle and another about Kripke. Searle considers a different sort of usage-contingency; instead of the point that the lexical meanings of expressions are usage-contingent, he suggests that the coreference of the two occurrences of 'Tully' in 'Tully is Tully' is usage--contingent, contingent on how we happen to use language (Searle 1958, 167). The difference between these two kinds of usage-contingency doesn't matter for my purposes.
     In talking about statements of English being metaphysically necessary, Kripke stresses that he is holding fixed that "we use *English* with *our* meanings and our references": "[o]ne doesn't say that 'two plus two equals four' is contingent because people might have spoken a language in which 'two plus two equals four' meant that seven is even" (Kripke (1970/1980, 77, emphasis in the original). So Kripke excludes both kinds of usage-contingency when discussing statements being necessary or contingent. That doesn't change the fact that it is a *metaphysical* possibility that linguistic rules are different (and this is what usage-contingency is about); it's just that this possibility is irrelevant when considering whether various statements of English, including 'Cicero is Tully' and 'Two plus two equals four' express necessary truths (according to Kripke, both of them do).

ous that the occurrences of 'necessary' and 'contingent' in the passage Kripke and others have focused on don't involve that notion.

The upshot of this section has been this: Searle's notion of necessity in his key passage was analyticity rather than metaphysical necessity; yet the unwanted necessity objection attributes to him the latter, thus missing its target. One might, however, worry: isn't there a related objection to be raised for Searle's own notion of necessity also? We might call it the *unwanted analyticity objection*: in the quoted passage, Searle is committed to (3) – "Aristotle did at least one of the deeds featured in the cluster" – being analytic, that is, true in virtue of linguistic rules, and hence knowable by a competent user of 'Aristotle' a priori, without recourse to experience. Kripke (1970/1980, 67-68) suggests that this is problematic, for it constitutes an empirical discovery (and not knowable a priori) about the bearer of 'Aristotle' that he did any of the deeds featured in the cluster associated with the name; indeed, occasionally, empirical discovery confirms the opposite, as in the case of Jonah, whom some historians consider a historical figure who did none of the deeds that the Bible attributes to him. These historians' position strikes us as perfectly coherent, yet Searle's passage commits him to the analyticity and thus the apriporicity of "Jonah did at least one of the deeds featured in the cluster" and hence to the historians' holding a contradictory stance.

In the next two sections, I will show that not only does Searle's own view – the presuppositional view – have the capacity to respond to objections (i) and (ii), but also to the unwanted analyticity objection.[10] Before returning to the objections, however, let's first lay out the presuppositional view.

4   The presuppositional view

What was Searle's own account, and just how different was it from the cluster view philosophers tend to attribute to him? Searle's description clusters turn out to play a markedly different role than that pos-

_____

[10]   In this paper, I don't assess the strength of the unwanted analyticity objection: instead, I grant it for the sake of argument and show that the presuppositional view can offer a response to it.

ited by the cluster view, which ignores the role of presuppositions in Searle's account.

What are the descriptions to be included in the cluster that gives the meaning of 'Aristotle'? Searle suggests that "we ask the users of the name 'Aristotle' to state what they regard as certain *essential and established facts* about him" (Searle 1958, 171, emphasis added). This does seem to invite the principled basis as well as the unwanted ambiguity objections. On what grounds do we separate what's essential and established from what isn't? And how do we avoid a situation in which what is essential and established varies from one individual to the next, thus varying the clusters and hence the meaning across speakers? Yet these lines of criticism ignore an essential detail about Searle's account.

To get a sense of Searle's view, we first need to recognize that he draws a crucial distinction between *extraordinary uses* of proper names and ordinary uses—*referring uses* as he calls them:

> … though proper names do not normally assert or specify any characteristics, their *referring uses* nonetheless *presuppose* that the object to which they purport to refer has certain characteristics. … Now what I am arguing is that the descriptive force of 'This is Aristotle' is to *assert* that a sufficient but so far unspecified number of these statements are true of this object. Therefore, *referring uses* of 'Aristotle' *presuppose* the existence of an object of whom a sufficient but so far unspecified number of these statements are true. *To use a proper name referringly is to presuppose the truth of certain uniquely referring descriptive statements, but it is not ordinarily to assert these statements* or even to indicate which exactly are presupposed. (Searle 1958, 170-171, emphases added)

Searle is quite explicit here: the referring uses of 'Aristotle' are such that they presuppose without asserting that the bearer of the name fits a sufficient number of the descriptions in the cluster for 'Aristotle'. (Crucially, the sufficient number of descriptions together have to identify Aristotle uniquely.) It is in the extraordinary cases (Searle mentions above 'This is Aristotle') that in making a statement, the speaker asserts (and doesn't merely presuppose) that the bearer of 'Aristotle' fits a sufficient number of the descriptions in the cluster. The extraordinary cases that Searle discusses include, on the one hand, *identity claims* ('This is Aristotle' as well as 'Tully is Cicero', 'Chomolungma is Mount Everest'), and on the other hand, *singular existential claims* like 'Aristotle never

existed', which, according to Searle, "asserts that a sufficient number of the conventional presuppositions, descriptive statements, or referring uses of 'Aristotle' are false" (Searle 1958, 173).[11]

It remains to be seen whether a suitable notion of presupposition would allow a Searlean view to avoid the principled basis, the unwanted ambiguity, and the unwanted analyticity objections for referring uses. The rest of this section will explore this issue. Another question, how the objections can be addressed in connection with the extraordinary cases, will be taken up in the next section.

What is the notion of presupposition plausibly at work in referring uses? There are two options to consider: in the case of a proper-name-containing sentence, the truth conditions of the utterance either feature (at least some of) the descriptions in the cluster, or they don't, featuring only the object to which the name refers, that is, the object uniquely picked out by the weighted most of the descriptions in the cluster. Let's say that in the former case, the utterance has *description-featuring truth conditions*; in the latter, *object-featuring truth conditions*. My aim is to show, first, that (a) Searle didn't take a clear stand on the choice between object- and description-featuring truth conditions, but (b) it is object-featuring truth conditions that are closer to his purposes, and (c) going object-featuring is a promising move for it affords an elegant response to extant objections for referring uses.

Let's begin with (a); Searle, I will argue, did not make clear whether he wants to construe the truth conditions of utterances involving proper names as description- or object-featuring. Let's first review considerations that speak *against* object-featuring truth conditions. Searle is explicitly citing Strawson's (1950) paper "On Referring" as his reference point for presuppositions: "Following Strawson we may say that referring uses of both proper names and definite descriptions presuppose

---

[11]   One might include among the special cases indirect discourse also, as in 'John believes that Tully was an orator'; Searle (1958, 1969) doesn't mention such cases.

On a side note, it is worth stressing that the range of referring uses of expressions is vastly broader than Donnellan's (1966) category of the so-called referential uses of definite descriptions; so the two labels should be carefully distinguished. For one thing, expressions other than definite descriptions have referring uses (proper names, for example); for another, the examples that Donnellan labels referential and attributive uses of definite descriptions, for example, the two uses of 'Smith's murderer is insane' (or: 'The guy who murdered Smith is insane'), *all* count as referring uses in Searle's sense.

the existence of one and only one object referred to" (Searle 1958, 170). And for the overall purposes of Strawson's paper, the choice of object-featuring truth conditions would be disastrous; here is why. Strawson's paper doesn't mention proper names at all; he discusses primarily definite descriptions in the context of utterances like the contemporary utterance of 'The king of France is wise'. Now, surely, it would be bizarre to suggest that this sentence, when uttered presently, has incomplete truth conditions due to there being no French monarch at present, and hence no object to be featured in the truth conditions of the utterance. It would be likewise bizarre to suggest that the truth conditions of the utterance do not feature descriptive material like being French or being a king. That would mean the following four sentences, uttered now, would have the very same object-featuring truth conditions:

(5) The Queen of England is wise.
(6) The successor of King George VI is wise.
(7) The only monarch ever to participate in a James Bond video clip is wise.
(8) The only monarch ever to participate in a Summer Olympics video clip is wise.

The object-featuring truth conditions for all of these utterances would involve a certain woman, Elisabeth II, presenting her as wise. By contrast, positing description-featuring truth conditions for (5)-(8) would give far more plausible candidates — they would allow for distinguishing among the truth conditions of the four utterances due to differences in the descriptive material featured in each.[12]

Elsewhere in "Proper Names", Searle himself doesn't take a stand on the choice between description-featuring and object-featuring truth conditions as his model for presuppositions. Beyond citing Strawson, he writes:

> But the uniqueness and immense pragmatic convenience of proper names in our language lie precisely in the fact that they enable us to refer publicly to objects without being forced to raise issues and come to agreement on what descriptive characteristics exactly con-

---

[12] Indeed, Soames (1989, 609, fn. 16) assumes that for (5)-(8), Strawson is opting for description-featuring truth conditions. Soames does not consider Strawson's subsequent (1959) commitments about presuppositions, to be discussed shortly.

stitute the identity of the object. They function not as descriptions but as pegs on which to hang descriptions. Thus the looseness of the criteria for proper names is a necessary condition for isolating the referring function from the describing function of proper names. (Searle 1958, 172)

The first sentence in the quote suggests that to refer with 'Aristotle', speakers need not come to an agreement about what descriptions identify him uniquely; But this can be accommodated with a description cluster (rather than a single meaning-giving description) for 'Aristotle', the weighted majority of which fit the bearer of the name and him only. How presupposed description clusters work, whether they issue in description- or object-featuring truth conditions, is an issue left wide open for all Searle has stated so far. The second, enigmatic sentence about names being "pegs on which to hang descriptions", might be taken to inspire a model of object/peg-featuring truth conditions that individual speakers reach via some or other uniquely fitting description; but there is practically no guidance or ground given by Searle here or elsewhere to steer us in this direction.[13] The third sentence can be understood in two ways: is Searle talking about a necessary condition for the referring function of *expressions quite generally*, or of *proper names only*? The first option seems at odds with Strawson's (1950) idea that typical examples of the referring function of language involve definite descriptions as they occur, for instance, in (5)-(8). Overwhelmingly often, in the case of referring uses of definite descriptions like 'the reigning Queen of England', there is no looseness of criteria, no looseness in the description at issue: the speaker talks about someone who meets the criterion of being the reigning Queen of England; yet Strawson does want to isolate the referring function of such definite descriptions from their describing function (Strawson 1950, 334-344). The second option yields no such conflict with Strawson's claims: if in positing looseness as a necessary condition for isolating the referring function from the describing one, Searle is talking about proper names only, then he could maintain (along with Strawson) that no such necessary condition ap-

---

[13] Names as pegs on which to hang descriptions is the perfect metaphor for D'Cruz's account (some aspects of which will be discussed later) according to which "an ordinary proper name is a mere placeholder for an arbitrary ordinary definite description true of the given individual" (D'Cruz 2000, 721).

plies to definite descriptions. But this would take further explanation and discussion of how Searle's take on referring uses relates to Strawson's; yet the above passage is the only place Searle mentions the referring function of language or of names. As it stands, the foregoing passage from Searle has puzzling aspects and doesn't in the end support the choice between description- and object-featuring truth conditions with respect to the descriptions that are presupposed.

Moving on to (b), why insist that object-featuring truth conditions fit Searle's purposes? First, the odd consequences described for examples (5)-(8) do arise for definite descriptions, but not for proper names, which are the exclusive focus of Searle's paper. Indeed, a popular position since John Stuart Mill has it that in uttering 'Cicero was an orator' and 'Tully was an orator', speakers express the very same object-featuring truth conditions. Second, Strawson's (1950) work discusses, besides referring uses of definite descriptions (for which, as we have seen, description-featuring truth conditions seem by far the more plausible of the two approaches), also context-sensitive expressions like 'I' and 'this'; and for sentences like 'I am a philosopher', 'This is red', positing object-featuring truth conditions is the vastly more plausible of the two approaches as David Kaplan argued since the 1970s, primarily in his (1977/1989) monograph "Demonstratives". It thus seems unfounded to regard Strawson (1950) as having given *univocal* support to one or the other type of truth condition.[14] Third, in his subsequent book *Individuals* (1959, 180-194), Strawson is rather explicitly opting for the object-featuring model of presuppositions: he suggests that a condition for "introducing a particular into a proposition" (see Strawson 1959, 180), making it part of what the speaker says (Strawson 1959, 182), is that the speaker be able to provide unique identification of the particular object, the uniquely identifying descriptions for proper names forming their "presupposition set" (Strawson 1959, 192). Crucially, for successful name use, the speaker and hearer both need to be able to provide unique identification of the particular object, but their ways of identifying can be different (Strawson 1959, 183).

And finally reaching (c): opting for object-featuring truth conditions in the context of the presuppositional view gives the crucial advantage

---

[14] See Soames' (1989, 562-566) illuminating discussion on how Strawson's (1950) view might be construed and related to Frege's (1892/1952) notion of presupposition.

of countering, in the ordinary cases, the remaining objections: about the lack of a principled basis, as well as about unwanted ambiguity and unwanted analyticity. By going object-featuring, the principled basis objection, (i), does not arise with respect to the truth conditions of utterances involving proper names: for the truth conditions feature the object only, so at that point, no principled basis is required for selecting among descriptions that are in the presupposition cluster, and ones that aren't. The unwanted ambiguity objection, (ii), does not arise either: again, the truth conditions for an utterance involving a proper name feature the object only for both speaker and hearer, despite their presupposition clusters being potentially different; at the level of truth conditions, 'Aristotle' makes the same contribution – the person – across language users. Further, the unwanted analyticity objection (which replaced the unwanted necessity objection once we clarified the right notion of necessity to attribute to Searle) doesn't pose a problem: in the case of referring uses of proper names as in (3) – "Aristotle did at least one of the deeds featured in the cluster" – the descriptions in the presupposition cluster are not featured in the truth conditions of the utterance, only the object is; this way, (3) isn't an analytic truth.

In sum, opting for object-featuring truth conditions for utterances involving proper names—that is, excluding the presupposed descriptions from the truth conditions of the utterance—is independently motivated and successfully responds to objections (i) and (ii) raised against the cluster view as well as to the unwanted analyticity objection. So far, we have covered only the ordinary cases, and it remains to be shown if the presuppositional view has the resources to handle extant objections with respect to the extraordinary cases; to this we now turn.

5   Deflecting what remains of the objections

Searle highlighted two kinds of extraordinary cases: identity claims and singular existential claims, suggesting that in such cases, there being a unique individual who fits the weighted most of the descriptions in the cluster is asserted and not merely presupposed. How might this part of the Searlean view be squared with objections (i) and (ii) – about a principled basis and unwanted ambiguities?[15] In this paper, I content

---

[15]   The remaining objection about unwanted analyticity doesn't arise in the extraordinary uses, so I will discuss it in footnotes only.

myself with developing a solution for singular existential claims, leaving identity claims to another occasion.

Quite independently of Searle's treatment of singular existential claims like 'Aristotle existed' or 'Aristotle never existed', or 'Aristotle didn't (really) exist', several philosophers have observed that such claims are distinctive in that they lack a *stable semantic role*.[16] Ludwig Wittgenstein writes:

> If one says 'Moses does not exist', this may mean various things.
> It may mean: the Israelites did not have a *single* leader when they
> withdrew from Egypt – or: their leader was not called Moses – or:
> etc. etc. – We may say, following Russell: the name 'Moses' can be
> defined by means of various descriptions … And according as we
> assume one definition or another the proposition 'Moses does not
> exist' acquires a different sense and so does every other proposition
> about Moses. (Wittgenstein 1953, sc. 79)

Wittgenstein is suggesting here that 'Moses doesn't exist' may variously mean 'the Israelites didn't have a single leader', 'The leader of the Israelites wasn't called Moses', and so on. When someone utters 'Moses doesn't exist', it is simply unclear what the truth conditions she expressed are; there are various nonequivalent candidates and no basis for choosing one over the others as the truth conditions of her utterance. It is only natural then that the utterance is ambiguous among speakers and that there isn't a principled basis for selecting which of the various things a speaker might mean by an utterance. Utterances of this sort are without a stable semantic role, it is to be expected then that objections (i) and (ii) should arise; we should be worried if they didn't.[17] This is a compelling line to take on singular existential claims. Wittgenstein does, however, in the last sentence of the passage above, generalize the point to *every* utterance involving the name 'Moses', including 'Moses had a beard'. On the one hand, the more general point lacks sufficient motivation: there is no expectation that 'Moses had a beard' lacks specific, stable truth conditions. On the other hand, this last point of Wittgenstein's provides evidence that he (unlike Searle)

---

[16] See D'Cruz's (2000, 740-743) thorough discussion.

[17] The unwanted analyticity objection does not arise for singular existential claims—clearly, statements of that form are never analytic on Searle's presuppositional account (nor are they analytic on the description view or the cluster view).

held the cluster view, against which objections (i) and (ii) were justifiably raised.[18]

But other authors like Gareth Evans (see Evans 1982, 396-398) and Mark D'Cruz (see D'Cruz 2000, 740-743) insisted, rightly, I think, that in the case of singular existential claims – more precisely, they highlight *negative* existential claims only – the phenomenon of there being no stable semantic role is quite robust, and deserves an explanation. Searle's proposed treatment for the extraordinary cases provides just this sort of explanation: given the cluster of descriptions being featured in what is asserted, the result is that we are faced with various descriptions and no principled basis to choose among them; we likewise expect ambiguity across speakers. When it comes to negative existential claims, objections (i) and (ii) arising against the presuppositional view is therefore something the proponent of the presuppositional view should consider an asset rather than a liability; it is those accounts that steer clear of (i) and (ii) that thereby face a disadvantage and need to explain their case.

Moreover, D'Cruz stresses the *contrast* between negative existential claims and other, more ordinary claims like 'Moses had a beard', which do have a stable semantic role:

> 'Aristotle is fat' has a stable semantic role in the language, in the sense that competent listeners would know exactly what to make of it without further ado. Thus, its utterance would not normally invite such remarks as 'What do you mean?' or 'I do not understand', which would belie its alleged stable role. Naturally, this stable role is derived from the stable role of its constituents – 'Aristotle', 'is' and 'fat' – and the way they are strung together in the utterance. An utterance of 'Q is fat', however has a stable role in exactly the same way, derived from the stable role of *its* constituents, and the way *they* are put together. Its utterance, therefore need not semantically puzzle one who already grasps 'Q' and who otherwise speaks English: he or she would know what to make of it. … Evans pointed out an interesting fact about a negative existential such as 'Ronald Reagan does not exist', involving a *mature* name-using practice, namely, that it has *no stable semantic role* in the sense just described … (D'Cruz 2000, 740, emphasis in the original)

_____

[18]   Indeed, besides Searle and Strawson, the cluster view is widely attributed to Wittgenstein also, based on this particular section, Section 79 of "Philosophical Investigations".

At this point, to give more detail about the absence of a stable semantic role in the case of negative existential claims, D'Cruz (2000, 740-741) goes on to quote Evans (1982, 397):

> When there is no specific body of information which is generally associated with a name (as in the case of a mature name-using practice), the negative existential statement does not have a clear sense … If someone said to me, 'Ronald Reagan does not exist', I should not know what to make of it. If the remark is intended to have a content such that it is true if and only if 'Ronald Reagan', as used in a certain name-using practice, does not refer, then I cannot conclude from its truth anything about the information I associate with the name, since for all that the remark, so understood, tells me, that information could still constitute knowledge about some individual.

D'Cruz is calling attention to this phenomenon: upon hearing 'Ronald Reagan doesn't exist', I cannot conclude from it anything about the definite descriptions 'the 40th President of the United States', 'the man who prior to becoming the 40th President of the United States had been an actor in movies like *Bedtime for Bonzo*, and had served as the 33rd Governor of California', and so on; for all I know, these descriptions might still fit some person or other.

A Searlean presuppositional view with object-featuring truth conditions receives substantial support from the contrast that D'Cruz is describing. There is something very intuitive about negative existential claims lacking a stable semantic role: we encounter semantic instability there, though not with the ordinary, referring uses of proper names as in 'Moses had a beard'. This contrast is entirely unsurprising by the presuppositional view's lights. Given what it takes for an individual to associate one or more descriptions with a name (she is to believe that the descriptions apply to the bearer of the name), the presuppositional view is – quite independently of D'Cruz's, Evans's and Wittgenstein's considerations – set up so in the ordinary cases, when the associated descriptions are merely presupposed, the variation among speaker associations does not interfere with the stable semantic function of the utterance, whereas in the extraordinary cases, when the associations become part of what is asserted, the result is messy and unstable.

Further, Evans's and D'Cruz's point that certain utterances (unlike others) are without a stable semantic role can and should, I take it, be generalized to *affirmative* existential claims like 'Moses existed', 'Homer existed', and 'Ronald Reagan existed' as well: these, too, are without

a stable semantic role.[19] Just as with negative existential claims, in affirmative ones (true and false claims alike), there are various things speakers might mean by them – a variety of nonequivalent truth conditions each of which are candidates for what the speaker meant – and no way for an audience to choose among them. D'Cruz's characterization of the contrast between the existence versus the absence of a stable semantic role (quoted above) applies equally to referring uses of proper names versus affirmative existential claims. Interpreting 'Moses existed' generates comparable puzzlement as 'Moses didn't exist' does.

To home in on just how robust the lack of a stable semantic role is in the case of negative existential claims, Evans draws a contrast between ordinary proper names and what he calls *thin uses* of proper names, which are strongly associated with a specific description: "I may book a flight in a false name, and then the next day telephone the airline and say 'Look, Agatha Hermer doesn't exist' … All that the receptionist need conclude is that when I uttered the name previously, I referred to nothing" (cf. Evans 1982, 398). The utterance 'Agatha Hermer doesn't exist' does have a stable semantic role. This is crucially unlike 'Moses didn't exist', 'Aristotle never existed' both featuring ordinary proper names and exhibiting semantic instability in need of explanation. Moreover, 'Moses existed', 'Aristotle existed', and 'Homer existed' likewise show semantic instability in need of explanation.

The lack of a stable semantic role in the case of existential claims is not explored by Searle, although he alludes to it in the last sentence of this passage:

> 'Aristotle never existed' … asserts that a sufficient number of the conventional presuppositions, descriptive statements, of referring uses of 'Aristotle' are false. Precisely which statements are asserted to be false is not yet clear, for what precise conditions constitute the criteria for applying 'Aristotle' is not yet laid down by the language. (Searle 1958, 73)

Here, Searle's point is that given the fact that the cluster of descriptions for 'Aristotle' does not provide a set of descriptions all of which are known to the speaker and are true of Aristotle, there is no single claim

---

19   Of course, these considerations about singular existential claims are intended to be quite general, covering (a) sentences in the present tense and (b) sentences involving proper names of things other than people, for example, 'Troy exists', 'Troy doesn't exist', 'Atlantis exists', 'Atlantis doesn't exist'.

of the form 'Nothing fits description D' that a speaker might be taken to mean by 'Aristotle never existed'. But Searle appears to raise this point more as a consequence of his view rather than as an independently motivated desideratum, which Wittgenstein, Evans and D'Cruz take it to be.

As Searle's account presently stands, the other type of extraordinary case, an identity claim like (4) (repeated below) also lacks a stable semantic role:

(4)   Tully is Cicero

It isn't settled exactly which descriptions are featured in the assertion when someone utters (4). Yet we don't have the kind of robust expectation of semantic instability that we had for singular existential claims. The identity claim simply doesn't invite the question: do you mean that the orator who spoke up against Mark Anthony is identical to … or do you mean that the statesman and philosopher whose name means chickpea is identical to …?, until that's clarified, it's unclear what you've meant. This suggests that on one minor point, with respect to identity claims, the principled basis and the unwanted ambiguity objections do have some traction against Searle's presuppositional account. Developing an alternative proposal for identity claims therefore remains an outstanding challenge.

I'd like to close this section by pointing out how this one aspect of deficiency concerning identity claims like (4) is connected to Searle's remarks about analyticity. Recall (from Section 3) that Searle suggested that (4) *can be used* to make an analytic statement and also a synthetic one (cf. Searle 1958, 167); he claimed also that (4) "is analytic for most people" (Searle 1958, 173). Searle thus seems to subscribe to a notion of analytic-for-an-individual rather than analytic-for-a-linguistic-community. Bear in mind, however, that the issue of a statement being analytic for some individuals but not others is quite limited on the presuppositional view we have developed: it isn't as though *referring uses* of names yield statements that are analytic for some individuals and not others. Variation among individuals arises only in extraordinary cases like (4) in which variation in the associated descriptions issues in variation in what's asserted, and hence variation in analytic/synthetic status.[20]

_____

[20]   Now we are in a position to address an issue raised in footnote 6: it turns out that on Searle's proposal, a single individual may use (4) to make a

Searle's individual-relativized notion of analyticity arising with respect to (4) does invite the two objections: what's a principled basis for selecting which description are and aren't in the cluster featured in the assertion?; and what do we do about unwanted variation in meaning across speakers?[21] The fact that Searle lacks a notion of analyticity for a linguistic community in connection with identity claims like (4) goes hand in hand with his view being susceptible to the two objections. The susceptibility is extremely limited, however: it concerns only one type of extraordinary case: identity claims.

6   Concluding remarks

Towards the beginning, I cited Searle's own summary of his view in "Proper Names": according to him, descriptive presuppositions provide a loose sort of logical connection between proper names and definite descriptions that fit the object that is the bearer of the name. In this paper, I have tried to flesh out this summary to show just how different Searle's presuppositional view is from the cluster view philosophers like Devitt – Sterelny (1999) and Ludlow (2007) attribute to him. These philosophers have concentrated on the "loose sort" aspect of Searle's characterization; they focused on clusters of descriptions rather than

––––––––––––––––––

synthetic statement at one time and an analytic statement at another (so it is more accurate to talk about Searle's use- or occasion-relative notion of analyticity). Let me explain. Recall that in extraordinary cases, the associated descriptions are part of what's asserted when making an utterance; now, the descriptions a speaker associates with a name can change over time (some descriptions are added, some removed or revised) *without the linguistic rules (the cluster for 'Tully' and 'Cicero', say) changing*; as a result, a speaker might at one time use (4) to make an analytic statement, and at another time, to make a synthetic one while the linguistic rules remain unaltered. Crucially, this sort of occasion-relative analyticity arises for the extraordinary cases only, in which descriptions associated by the individual make it into what's asserted.

[21] The remaining objection, about unwanted analyticity, doesn't arise in the context of Searle's account of identity statements: the only way it could arise is if we want 'Cicero is Tully' to *not* be analytic for a certain individual yet Searle's presuppositional view would make the statement come out analytic. But in such a situation, with the individual's descriptions associated with 'Cicero' and 'Tully' being different and being featured in what is asserted when making the identity claim, the result wouldn't be an analytic statement.

a single description giving the meaning of a proper name. But they ignored the presupposition aspect that I aimed to bring to the fore, showing that it yields an interesting view worth reconsidering. Searle's presuppositional view has two parts. First, in *ordinary cases* — referring uses of proper names that include most uses apart from the exceptions to be cited — utterances involving 'Aristotle' *presuppose* that the weighted majority of the cluster for 'Aristotle' fits the bearer of the name. Second, in the *extraordinary cases* — for which Searle cites just two types, singular existential claims and identity claims — utterances involving 'Aristotle' *assert* that the weighted majority of the cluster for 'Aristotle' fits the bearer of the name.

In the ordinary cases, the role of the description cluster is quite limited; I have argued that it is plausible and promising to develop Searle's view in a way that is consistent with his "Proper Names": my version of the presuppositional view denies that the descriptions in the cluster are featured at all in the truth conditions of the utterance; instead, the truth conditions plausibly feature the object only, to wit, the person Aristotle. This Searlean view is far closer to Kaplan's influential (1977/1989) post-Kripkean direct reference theory – according to which the only truth-conditional contribution of a proper name is its referent – than it was previously assumed. The idea that Searle's theory is a largely Fregean interlude between Frege's description theory of proper names and Kripke's proposals presented in "Naming and Necessity" is in major ways a myth, a mythical chapter in how the story of 20th-century philosophy of language is often told.

Granted: in the extraordinary cases, the description clusters step in to play a greater role than in the ordinary cases. Descriptions from the cluster for 'Aristotle' are part of what is asserted by an utterance like 'Aristotle never existed'. But it is well to bear in mind that the extraordinary cases are quite isolated, the exception rather than the norm.

The way philosophers have been telling the history of 20th-century philosophy of language prominently included three objections taken to apply to Searle's view: the principled basis, the unwanted ambiguity and the unwanted necessity objections. My aim has been to show that these objections leave Searle's presuppositional view largely untouched; moreover, the presuppositional view is an interesting one worthy of further consideration. I haven't tried to defend the presuppositional view against all Kripkean objections, focusing instead on just these three. In particular, the unwanted necessity objection mistakenly

attributes to Searle — writing "Proper Names" in the 1950s — Kripke's metaphysical notion of necessity, which did not emerge in a clarified form until at least a decade later.

Kripke was characteristically cautious, paying attention to the details of Searle's view more closely than Devitt, Sterelny and Ludlow did: he distinguished theories of meaning from theories of reference determination, and unlike the other philosophers, did not take it as a given that Searle was putting forth a theory of meaning in terms of description clusters (cf. Kripke 1970/1980, 31-34, 57-61).[22] Nonetheless, he did point out several times that a theory of reference determination is rather limited in scope: it doesn't provide a way to *analyze* any sentences involving proper names, including 'Aristotle never existed'. To give an analysis in terms of description clusters, what is needed is a theory of the meaning of proper names, Kripke pointed out in (1970/1980, 33-34, 58-59). From this, his audience and readers probably drew the conclusion that Searle's goal was to appeal to description clusters within a more ambitious theory of the *meaning* of proper names: the cluster theory. After all, Searle did say that 'Aristotle never existed' "asserts that a sufficient number of the conventional presuppositions, descriptive statements, of referring uses of 'Aristotle' are false" (Searle 1958, 173). But drawing this conclusion ignores the fact that Searle considered sin-

_____

[22] Kripke (1970/1980, 31, emphasis added) quotes a passage from Wittgenstein, following it up with "According to this view, and a *locus classicus* of it is Searle's article on proper names, the *referent of the name is determined* not by a single description but by some cluster or family". Kripke thus initially takes a conservative approach, interpreting Searle as giving a theory of reference determination. By contrast, Sterelny – Kim (1999, 50) and Ludlow (2007, Section 8), respectively, introduce Searle's (1958) view as follows:

> … the most influential exponents of the ["cluster" or modern] theory were Peter Strawson and John Searle. Instead of tying a name tightly to one definite description, as the classical theory goes, the modern theory ties it loosely to many. *This cluster of descriptions expresses the sense of the name and determines its reference*… (emphasis added)

> Consider a name like 'Socrates'. Is it really part of the meaning of that name that its bearer drank hemlock, taught Plato and did all the other things that we are told that he did when we study the history of philosophy? Searle suggests that we needn't associate the meaning of a name with a description that contains *all* of these elements — it might be enough if most of them hold, or that a suitably weighted bundle of them hold. (first emphasis added, second in the original)

gular existential claims to be out of the ordinary; he took the role of description clusters to be crucially different in the ordinary cases.

More generally, the very idea of offering just the two options – theory of meaning versus theory of reference determination – for characterizing Searle's view ignores two aspects of Searle's proposal: first, that he sets apart his treatment of the extraordinary and the ordinary cases, and second, that he employs the notion of presupposition in the latter cases. In the ordinary cases, neither option gives an accurate characterization of Searle's proposal: the description cluster for 'Aristotle' certainly isn't what gives the meaning of 'Aristotle' the way we usually understand what meanings are about; but nor is its role nothing over and above determining the reference of 'Aristotle'. This role presuppositions play in Searle's view has been widely ignored for half a century, giving rise, within an important chapter of the philosophy of language, to myth rather than history about Searle's (1958) "Proper Names".[23]

Department of Logic
Institute of Philosophy
Faculty of Humanities
Eötvös University
Múzeum krt. 4/i
Budapest, H-1088
Hungary
zvolenszky@elte.hu

References

Burgess, J. (1997): Quinus ab Omni Nævo Vindicatus. In: Kazmi, A. A. (ed.): *Meaning and Reference: Canadian Journal of Philosophy Supplement* 23, 25-65.

D'Cruz, M. (2000): A Theory of Ordinary Proper Names. *Mind* 109, 721-756.

Devitt, M. – Sterelny, M. (1999): *Language and Reality: An Introduction to the Philosophy of Language*. Cambridge, (Mass.): MIT.

Donnellan, K. (1966): Reference and Definite Descriptions. *Philosophical Review* 77, 281-304.

_____

EVANS, G. (1973): The Causal Theory of Names. *Proceedings of the Aristotelian Society*. Supplementary Volume 47, 187-208.

EVANS, G. (1982): *The Varieties of Reference*. McDowell, J. (ed.). Oxford: Clarendon.

FREGE, G. (1892/1952): On Sense and Meaning. In: Geach. P. – Black, M. (eds.): *Translations from the Philosophical Writings of Gottlob Frege*. Oxford: Blackwell, 56-78. Originally published in 1892 as Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik* 100, 25-50.

KAPLAN, D. (1977/1989): Demonstratives. In: Almog, J. – Perry, J. – Wettstein, H. (eds.): *Themes from Kaplan*. Oxford: Oxford, 1989, 481-563.

KRIPKE, S. (1970/1980): *Naming and Necessity*. Cambridge, (Mass.): Harvard.

LUDLOW, P. (2007): Descriptions. In: Zalta, E. (ed.): *The Stanford Encyclopedia of Philosophy* (Winter 2011 Edition), Available: http://plato.stanford.edu /archives /win11 /entries /descriptions

MARCUS, R. B. (1961): Modalities and Intensional Languages. *Synthese* 13, 303-322.

MARCUS, R. B. – QUINE, W. V. O. et al. (1962): Discussion. In: Wartofsky, M. (ed.): *Boston Studies in the Philosophy of Science*. Dordrecht: D. Reidel, 105-106.

NEALE, S. R. (2000): On a Milestone of Empiricism. In: Kotatko, P. – Orenstein, A. (eds.): *Knowledge, Language and Logic: Questions for Quine*. Dordrecht: Kluwer, 237-346.

QUINE, W. V. O. (1943): Notes on Existence and Necessity. *Journal of Philosophy* 40, 113-127.

QUINE, W. V. O. (1953): Reference and Modality. In: Quine, W. V. O.: *From a Logical Point of View*. Cambridge, (Mass.): Harvard, 139-159.

SEARLE, J. (1958): Proper Names. *Mind* 67, 166-173.

SEARLE, J. (1969): *Speech Acts*. Cambridge: University Printing House.

SOAMES, S. (1989): Presuppositions. In: Gabbay, D. – Guenthner, F. (eds.): *Handbook of Philosophical Logic*. Volume IV. Dordrecht: D. Reidel, 553-616.

SOAMES, S. (2002): *Beyond Rigidity: The Unifinished Semantic Agenda of Naming and Necessity*. Oxford: Oxford University Press.

STRAWSON, P. (1950): On Referring. *Mind* 59, 320-344.

STRAWSON, P. (1959): *Individuals*. New York: Routledge.

WITTGENSTEIN, L. (1953/1968): *Philosophical Investigations*. Translated by Anscombe, G. E. M. New York: Macmillan.

Proper Names and Their Role in Social Ontology

Marek Nagy

Palacký University, Olomouc

Abstract: The article deals with an applicability of John Searle's theory of social ontology to linguistic category of proper names. We suggest that in the context of Searle's theory we can consider proper names to be a kind of social institution. By several examples from anthropolinguistic research and from the field of theoretical onomastics we try to show that proper names meet conditions specified by John Searle, in particular that – through different communities all over the world – they are "holders" of various types of deontic power. At the end of our article we shortly weigh the possibility that proper names can be regard as status indicators (in terms of Searle's theory) too.

Keywords: baptism, deontic power, onomastics, proper names, social ontology.

> *Blake, John? Nothing here.*
> *Alright, try my legal name.*
> *You should use your full name.*
> *I like that name, Robin.*
> Dark Knight Rises

To speak about problems of proper names in the context of philosophical views of John Searle makes sense for various reasons. Primarily, professor Searle has dealt with the category of proper names repeatedly: In fact, he started his career with the text *Proper Names* (see Searle 1958) and the conception presented here brought him directly in "textbooks" of this important philosophical debate; proper names as rather strange and interesting individuals are aptly shown also in Searle's Speech Acts (see Searle 1969; the evidence of strangeness could be the mentioned and everlasting discussion on their reference behavior,

sense, meaning, denotation, connotation and differently called chains …). And so it may not be uninteresting to stop and think about the position these special entities (he dealt with repeatedly) could take in the concept of social ontology. It has been developed by professor Searle in the last decades. He speaks about determining position of natural language in the broader context of human acting and human society.

I will keep to a few short comments. They will refer to social-ontological dimension of proper names. I hope the connection will not be only accidental.

1   Social ontology concept

First, several words about the idea of social ontology. Professor Searle presented it in detail in the book *The Construction of Social Reality* (1995), he summed up its basic sources again e.g. in the article *Social Ontology: Some basic principles* (2006; it is our starting point in particular) and he developed it again in the book *Making the Social World* (2010). However, basic elements of the concept of social ontology and institutional facts, which professor Searle works out in the above mentioned works, are included in the earlier "language" oriented works (compare e.g. Searle 1965 and in it mentioned differentiation of regulative and constitutive rules or the Chapter 2.7 in *Speech Acts* devoted to differentiation of pure and institutional facts).

The central proposition of the concept reads: There is something that could be called social reality. And it exists only because we think it exists. At the same time professor Searle develops a very interesting conviction that it is the human language that is the necessary condition of existence of something like institutional facts constituting this social reality. The existence of these facts is set up by means of phenomenon of collective intentionality which enables attaching functions (status functions) connected with the key term of deontic power. The whole construction is summed up by professor Searle in the formula having the character of constitutive rules (his older concept), which is (by the way) an example representative of semiotic thinking: X constitutes Y in context Z. Professor Searle says convincingly that what we call society is an extremely complex net of those "status functions" constituting our social reality, namely on deontic basis of "positive and negative power". Professor Searle simultaneously stresses that it is the question of cases where deontic power is in play, i.e. the fact constituted as a re-

sult of collective intentionality attached to status Y in context Z (physical existence of X, as he shows in the example of limited company, is not always necessary) connected with the set of rights and obligations. From the point of view of philosophy of language, it is essential that the condition of existence of such (humanely specific) system is existence of language as a medium of representation. Professor Searle claims (but does not develop in detail) we need for it a sufficiently rich symbolic system (which language represents for him), it is even necessary because key status functions are formed by speech acts – declaratives. Making such speech acts, one of the components of social ontology – institutional fact is constituted. What professor Searle is interested in (besides the very fascinating development of the idea of "logical" bases of human society) are the types of institutional facts. We will keep to them.

Key terms of the conception of professor Searle are *status function*, *collective intentionality*, *deontic powers* a *desire-independent reasons*, which form special characteristics of human socialization. And in professor Searle's opinion, language, of course, which is a medium of representation. It is worth mentioning one important comment – in the concept of social ontology we speak about language in a broad sense. So the condition is not full-blown natural language, but simply a symbolic system which is an exclusive bearer of deontic power. The last comment is notable because it can cause doubts – Does it make sense in this context to deal with specific linguistic categories such as proper names? (When professor Searle doubts in his article if it depends whether language has this or that category and speaks about language "in a broad sense"). I think yes. In my opinion, proper names are a very specific and important social institution.

2   Professor Searle on proper names

As for proper names, professor Searle speaks about them both in *Proper Names* and in *Speech Acts* and he pays attention to them wholly in discourse of debates about their meaning and reference qualities. Proper names are for him a remarkable category. It is shown in his formulations such as "Yes, in a loose sort of way" (Searle 1969, 170; if proper names make sense), or "seldom we consider proper names as part of one language as opposed to another at all" (Searle 1969, 169; in margo of Frege's example with dr. Lauben and the declaration that those who

connect with certain names different descriptions speak different languages). Such statements indicate the status of proper names is rather ambivalent. Professor Searle used in both his texts a funny metaphor of proper names as something like hangers (professor Searle calls them "pegs") for description. People sometimes have to hang something on them so that hangers can be useful (and agree somehow what is hanging on them), but on the other hand, they must not forget that clothes on hangers are something else than clothes in a heap (we know it well from our households). How I think these hangers exist will follow in a minute.

3   Proper names as an institution sui generis

As for proper names, fortunately, we have a special linguistic discipline called onomastics at our disposal. The field of onomastics is complex: beside the care of semantics and reference of proper names – it shares this with logic and philosophy of language – it also includes other linguistic matter (e.g. parts of speech, syntactic and word-formative characteristics of proper names) and for quite a long time it also studies sociological aspects of proper names (in narrow interconnection with ethnology, for example; remember Frazer's *Golden Bough*). No wonder there is a lot to say about proper names from sociolinguistic point (from socially semantic fields of proper names to the phenomenon of family relationship expressed by surnames) – proper names rank among so-called language universals (cf. Trost 1995a). It seems there is an agreement that every language community has in its repertoire expressions which could be classified as proper names (to be honest, sometimes it is a tricky stuff because some proper names look much more like descriptions, but so do Morning Star and Evening Star anyway). In the same way it goes that comparing various societies and their handling proper names reveals a great spectrum of special characteristics of sociolinguistic nature. One of the basic warnings definitely remains (says Paul Ziff to Saul Kripke in Ziff 1977), to come with some generally valid theory of proper names is unusually daring (undoubtedly in what I am going to speak about there will be difference between the names of persons and hills, namely in different communities, not speaking about such proper names as FC Liverpool; I will only speak about personal proper names, so-called anthroponyms). "Any 'picture' of proper names that ignores such data that ignores the evolutionary diachronic

character of names and the languages in which names are used is not worth hanging" (Ziff 1977, 332).

In spite of my own warning, I would like to make some more general comments on proper name as a social institution. I will have to mention such things as baptism, identity, index and state of health. I am afraid even in this case the nature of proper names will appear slightly mysterious.

Let us remind again what are, by professor Searle, key moments of forming institutional facts by means of attaching status function: they are language (symbolic system of representation) as an instrument of forming a fact (declarative), collective intentionality (collective acceptance) and deontic power/character of fact (its connection with the set of positive/negative rights/obligations).

Let us start ab ovo, with what Saul Kripke calls "baptism". How does it happen? Undoubtedly, a part of it is some speech act. Professor Searle mentions in his texts in connection with institutional facts forms of declarative as a key speech act. Declarative, as he says, leads from words to the world and vice versa at the same time (in contrast with directive = words → state of world or constative = words ← state of world). Such declarative is naming someone a king or founding a limited company (I trust professor Searle that at least by California law it is like this). The very speech act of declarative establishes an institutional fact (e.g. real formation of legal person). Then could we consider baptism a speech act characterized as declarative? I suppose so. And I find support in one of older texts of professor Searle (see his 1976). Baptism goes as follows: under the given conditions enabling the community subsequent acceptance and use of proper name, semiotic relation between the subject as individual and the name of this individual is established. In Wittgenstein's opinion, the case of name is not the question of move in a language game (that is what I agree with – the name was not used but mentioned), but the very baptism has the character of speech act in relation to the subject and the community and it is a move in a language game, in my opinion.

I say baptism, but naturally I mean any moment when proper name is given in the procedure which various communities and cultures consider appropriate. Of course, I can name anything in any way (and as an atheist, thank God). But language is – as de Saussure or Central European interwar structuralists taught us – a fundamentally social phenomenon where nothing much happens without censorship/

acceptance by language community. And there are numerous proofs showing that for cultures of the world this act is an extremely serious moment tied up with socially obligatory rules so that it could be accepted by community. Probably most communities have regulated conditions for introducing names. And it does not matter if they are Wishram with their ritual specialist who is the only one authorized to choose a child proper name, or Czech parents at City Office in Hradec Králové who choose it themselves, but they have to do it and confirm it with their signature (i.e. with their proper name in the individualized format; they have right to do it if they introduce their proper name and produce evidence with the document containing the name!). In tribe Ga when an infant is given a name, members of all four branches of his kindred must be present. Delaware child's name is first spoken to the Creator and then repeated reverently so that the Creator will remember the child by name. And so on. No wonder, across communities of the whole world, hardly any act of behavior is so strictly determined with the net of social rules as the act of attaching proper name to its bearer – it is an important creative event (very ceremonial – Catholic christening, or technically mechanical – visit in a social department of a local office), even if with various implications. What is common to these events is that by the act of baptism, by introducing name its new bearer is socially individualized inside the given community, i.e. he enters it as a RIGHTful individual. "In our society there are no nameless, everyone has not only right to have a two-part name but it is his duty to have it. Conditions are regulated e.g. by Register Law. It does not apply only to our country: "In American society a personal name is attached to its bearer by law such that a name change must be legally notarized, and one' s signature is used to make agreements legally binding"" (Bean 1980, 311).

The act of giving proper name offers an individual the right to be regarded a member of the community. In a loose "Quinean" metaphor it is a dummy variable which is evaluated by baptism and in this way socially ontological obligation of existence is met. In baptism community gives rights to an individual and simultaneously it accepts culturally various sets of obligations. The importance of community acceptance and its indisputability is naturally basic and obvious. In some cultures in the act of baptism all the community members have to repeat the given name which confirms the acceptance of the name and the individual.

Pavel Trost in (1995a) characterizes this process of awarding and "owning" a name as the foundation of theoretical identity of the subject, which further of socially constitutes the given individual. "Further, while both proper names and definite descriptions identify individuals, the indexical character of proper names, that is their connection to their bearers, makes proper names not simply descriptive of the individuality of their bearers, but constitutive of it as well. A proper name is part of the individual identity of its bearer whether the latter be a person, a mountain, a river. (It may be because of this pragmatic linguistic fact that in so many societies personal names are considered to be part of the self or soul and naming constitutive of social persona)" (Bean 1980, 308). This identity, with only minor exceptions, does not change while other features of the subject change. Name may be the only one which survives. (Again even here we can find social difference and systematics – e.g. there are tribes in which the change of name is connected with social recognition of personality transformation and with the change of social status). Many mythologies of the world confirm this tight connection of proper name with bearer´s identity and its constitutive role (in tribe Bantu the given name determines the character of its bearer – it binds him to a certain type of social behavior). One of the manifests EZLN (Ejercito Zapatista de Liberacion Nacional) says: "Now we are named, we cannot die." I consider this an apt intuition. As if the nameless could not ask for their social rights, as if they did not even exist. Article 10 of Basic Document of Rights and Liberties says everyone has the right to have his name protected. In tens of law regulations we can find formulations concerning the fact if someone acts or does not act in his proper name.[1] It is interesting that the right to personal proper name used to be – and somewhere still may be – limited. One needs only to remember the practice of treating proper names in Ancient Greece and Rome. "Personal name is a social moral value; anonymity is a shame" (Trost 1995c, 313).

Proper name is something like an identity anchor not only from society point of view – it protects identity against changes of descriptions of subject – but even from point of view of its bearer. It is not by chance – as an Marek Tomeček's speech *The Name in the work On Certainty*[2]

---

[1] See References.

[2] Speech held during seminary "Wittgenstein" which took place in Plzeň in 2009.

reminded me – proper names serve as a diagnosis tool with patients after recovering from unconsciousness. If a patient can't remember his name, something very bad happened. In cinemas we can enjoy the fourth part of agent Bourne's story. The story about a man who had dozens of passports with tens of names in them and out of blue he can't remember which one is the real one. He can't remember his name and that's one of the symptoms of his identity loss. And not only inner one (amnesia, impossibility to reconstruct his personality through a name as an identity anchor), but even outer – he is torn out from the web of social, legal but even communication-pragmatic relationships. Who will he be when he cannot confirm his identity, sign a contract, fill in forms? In which sense is he still a member of our society? Proper name connects us socially to our actions, to our past. This is the very socially-existential reason why we can talk about institution of proper names fulfilling deontic power condition. It founds obligations and rights. It itself is right and obligation.

I know, it may look weird or even like manifestation of primitive thinking (which is ironized by Frazer who himself is reproached by Wittgenstein for the same thing for a change): all that magic and belief which say that a name is a part of soul, that a name is something more than useful means of reference, that to know somebody's (true) name is to have control of him. Taboo of a real proper name. All this sounds suspicious. Newspaper readers in Czech were prepared for a dull season recently by cause of a dumb foundling. He had no name by the way (more precisely he did not say it to the police). Can we say that the nameless boy lost his right to legitimate trial, medical care etc.? Absolutely not – evidently he is a human being and he would use all of this even if he wasn't ever baptized and lived his life somewhere in the jungle like Mowgli.[3] But frankly – could he, later, buy a house, run business, ask for child benefits being still nameless? Probably not. The fact that he has a name opens all these possibilities on the other side (precisely it is a necessary condition not a sufficient one – we are talking about net of connected status functions here).

In this context professor Searle mentions several ways of creating institutional facts. Is baptism creatio ex nihilo (precisely from a net of status functions)? I think this is not the case. There is a physical entity

---

[3] In fact Mowgli evidently has his name. But he was baptized by his adoptive wolf mother in the language of animals which people cannot understand.

which, as many of us believe, is an individual (it has a unique mind and DNA), having a name or not (but maybe someone doesn't think this way, see unity of name and soul in Inuite mythology). It is the case "X counts as Y in C and We accept (S has power (S does A))" (see Searle 2006, 65) more likely than the case "We make it case by Declaration that an entity Y exists that has status function(s) F in C" (cf. Searle, 2010, 100).

In context of professor Searle's theory I worry about another thing but I cannot deal with it now. Long story short. I took a stand that we can treat proper name itself as some kind of social institution. By means of specific speech act they create social objects – theoretical identity of subject, they are linguistic representations of status function which is created by its very existence. But there is some but in professor Searle's concept: he also mentions institution of something called "status indicator" (Searle 2006, 63). Things (!) that indicate (confirm, refer to) the fact that the given person is authorized or certified to some "status function". Some communities – discourses more likely – just insist on it. We are talking about passports, college diplomas etc. In connection with professor Searle's belief that proper names do not belong to "one language" and that we do not learn them in the way we learn other expressions (their meaning) suggests this interpretation to me: proper names stand somewhere at the edge of language, metaphorically speaking: as if they lean out of the language into the world of things, which proper names stand for in semiosis. What I'd like to say. This jazzy speculation about proper names as symptoms is maybe supported by Ch. S. Peirce's works. He labels proper name as an index or more precisely he classifies it as a subindex. The relation to what it stands for differs from the one represented by symbols, typical linguistic expressions. While symbols represent on the basis of general law and tradition, indexes "represent a comparison, a real connection, a clash, that we can associate with what happens in acts of volition or, more generally, of existence. Index has a force but neither sense nor character" (Peirce 1931-1935, 3.434). It looks as if proper names were names (language expressions) and symptoms of things at the same time. The name we have distinguishes us from others, it works as a distinctive feature (you can literally owe your name in some communities). Speech act of introducing is not a description but an action, performative. Czech linguist Pavel Trost in several papers says: proper name is literally a part of an object, it's not only a sign (signum), but even a thing (res). "Boy František differs from

the other boys in various psychophysical attributes, but also in the fact his name is František. This name doesn't signify his essential qualities, but it itself is an essential quality" (Trost 1995d, 258).

Maybe we can regard proper names as these status indicators in question (in cases, when we think of them as confirmation of status seen as a complex of rights and duties). "If this is your name, you must be a noble man." "Identify yourself – what is your name?" "My name is Fox Mulder. OK, you can enter then!" They would be an ID card, symptom of status role assigned to the subject.

4   Openendedness

This article doesn't supply argumentation in detail, it only suggests the ways we can try to think the problem through. They lead to the recognition of proper names as a specific part of social ontology. We suggest treating the very institute of proper name as a type of social institution. "To have a name" is a social concept and particular proper names are institutional facts. The very fact of using language as a sign is a socially significant performance and it supports the idea that natural language is a cornerstone of socialization.

Department of General Linguistics
Faculty of Arts
Palacký University
Křížkovského 10
771 80 Olomouc
Czech Republic
marek.nagy@upol.cz

References

BEAN, S. S. (1980): Ethnology and the Study of Proper Names. *Anthropological Linguistics* 22, 7, 305-316.
Česká republika. Listina základních práv a svobod. In: *1992*. 1992, č. 2. Dostupné z:http://www.psp.cz/docs/laws/listina.html
PEIRCE, Ch. S. (1931-1935): *Collected Papers. Volumes I–VI.* Cambridge: Harvard University Press.
SEARLE, J. (1958): Proper Names. *Mind* 67, 166-173.
SEARLE, J. (1965): What is a Speech Act? In: Black, M. (ed.): *Philosophy in America.* George Allen and Unwin, 221-239.

SEARLE, J. (1969): *Speech Acts. An Essay in the Philosophy of Language.* Cambridge: Cambridge University Press.

SEARLE, J. (1976): A Classification of Illocutionary Acts. *Language in Society* 5, 1, 1-23.

SEARLE, J. (2006): Social Ontology: Some Basic Principles. *Papers* 80, 65.

SEARLE, J. (2010): *Making the Social World.* Oxford: Oxford Univesity Press.

TROST, P. (1995a): Vlastní jméno-jazyková univerzálie? In: *Studie o jazycích a literatuře.* Praha: TORST, 293-294.

TROST, P. (1995b): Poznámky k teorii vlastního jména. In: *Studie o jazycích a literatuře.* Praha: TORST, 271-276.

TROST, P. (1995c): Ještě jednou o povaze vlastního jména. In: *Studie o jazycích a literatuře.* Praha: TORST, 309-313.

TROST, P. (1995d): Unika a vlastní jméno. In: *Studie o jazycích a literatuře.* Praha: TORST, 258-259.

ZIFF, P. (1977): About Proper Names. *Mind* 86, 319-332.

John Searle's Theory of Sign

Vít Gvoždiak

Univerzita Palackého, Olomouc

Abstract: The article attempts to link John Searle's philosophy and the area that is traditionally called semiotics, to bridge these domains and to demonstrate that they do relate to a shared bunch of problems. A brief discussion about the basic semiotic terms suggests that Searle's philosophy offers an explanatory framework to key semiotic questions, namely the differentiation of non-signs and signs, the place of intentionality in semiotic description, and the nature of sign correlations. As a consequence, Searle's theory can be called communication-oriented semiotics, which in the light of classical concepts developed by Peirce and de Saussure can be seen as a non-trivial contribution to the semiotic research.

Keywords: arbitrariness, intentionality, Peirce Charles S., representation, de Saussure Ferdinand, Searle John, semiotics, sign function.

There is a strange chasm between what is traditionally called semiotics and the work of John Searle. Very little, for instance, has been written on John Searle's theory in semiotic encyclopedic dictionaries (such as Sebeok 1986 or Cobley 2010), with the entries mostly restricted to relatively simple definitions of the speech act. Likewise, we can find hardly any mentions of semiotics, semiosis or the term sign in Searle's books. There are some clues (such as the term symbol or symbolize) that suggested that we could read, at least as an experiment, Searle's philosophy as an analog to semiotics.

The aim of this article, therefore, is simple, namely to bridge these domains and to demonstrate that they do relate to a shared bunch of problems and questions.

To start with, there are some reasonable objections as to why building or strengthening connections between Searle and semiotics be a needless effort that had better stop immediately. These objections, from

my point of view, are mainly terminological. In his article "Chomsky's Revolution in Linguistics", Searle (1972) discusses Chomsky's generative grammar as a milestone in the study of language and concludes that this milestone should be understood as a study of syntactical nature of language, which refuses to incorporate speech act theory, i.e. it does not offer any appropriate theory of semantic competence.

Similarly, we could define semiotics as a formal science (or doctrine) studying various relations between objects called signs with an emphasis on systematic and abstract nature of these objects and relations without much interest in the real-world or everyday communication taking place between real people. If this is to be the definition of semiotics, then such comparative effort is truly pointless. Unsurprisingly enough, there are nevertheless various points of view as to the precise manner and criteria under which semiotics is to be defined; therefore, should a particular definition prove inconvenient, we can always find another one, better suited to our needs. I will suggest one such definition shortly.

In general, the semiotic project can be characterized as an assemblage of topics introduced in its modern form by Charles Peirce and Ferdinand de Saussure and further developed by a plethora of their followers. Although there is hardly a reason or space to summarize any of the theories by these intellectual luminaries here, it is still somehow possible to derive the basic semiotic concerns from their work.

My aim is to reduce (and there will be no surprise there) the whole of the semiotic inquiry to a single concept, the sign, more specifically several selected theses elaborated on by de Saussure and Peirce. By adopting this scope, we can say that signs are (1) systematically arranged, (2) arbitrary and (3) social. Thus, if it is possible to speak about Searle's semiotic theory at all, it will be necessary to find some sort of correspondence between the aforementioned semiotic theses and some of Searle's terms. The most fundamental of these takes the form of "X counts as Y in context C" (see Searle 1969, 51-52 and 1995, 44) and can without hesitation be seen as a model of sign. Now, allow me to elaborate.

1   Functions

If we are to recognize certain things as signs, we simply have to have the ability to recognize them as signs. This mechanism can be

called the pragmatic rule, and was proposed by Charles Morris (see Morris 1938, 35). It can also be formulated using Searle's terms: signs are observer-relative (see Searle 1995, 12-13) and self-referential (see Searle 2010, 138), that is, no one can "fool all the people all the time" (see Searle 1995, 32) when it comes to such sign recognition. The pragmatic rule suggests that if we recognize something as belonging to a certain type we are dealing with a sign. This view however is not, semiotically speaking, precise, because recognizing things (trees, chairs, and even cocktail parties) as tokens of certain types does not necessarily imply that we understand them as signs.

What we face here are difficulties of various possible origins (individual, biological, cultural, etc.) and could be labeled epistemic. An illustration of these difficulties but also of the pragmatic rule is Ryle's (2009) example of two boys rapidly contracting their respective right eyelid. One of the boys is merely experiencing a tic, whereas the other one is winking at his friend with some conspiratorial intent. At first sight, the outcomes are identical, both tokens of the same type (contracting eyelid), but the difference lies in that while the first case is a matter of neurological causality, the second one is based on semiotic cooperation. The wink is a sign (X counts as Y in C) but the twitch is not (it is solely X). Thus the first semiotic finding is that signs are never intrinsic to our physical world (see Searle 2010, 14). Semiotics should be seen more exactly as social semiotics even though we can find some evidence for a somewhat similar mechanism in a medical check-up of the twitching boy concluding that his contracting eyelid is a symptom of some neural issue.

I think that we can get rid of such problems by introducing a definition of semiotics which can be stated as follows: semiotics is a study of every possible thing that can be used for lying (see Eco 1979, 6-7). We cannot lie using merely X, whereas we can lie or cheat (but also say the truth) with an "X as Y in C" device. Thus the nature of (sign) function cannot be reduced to causality of the physical world, that is, it exists in the triadic form XYZ unlike the brute, physical facts which are governed by physical laws (see Searle 2010, 10).

Without recognized sign functions we see only inevitable causal motions of a man crossing a line with a ball, whereas the sign function provides us with a notion of scoring and winning.

Albeit there is a significant confusion with regard to it, an X term in and of itself is not a sign but merely a physical signal which can "stand

for something else", yet this "something else" is not a Y term but rather another X term. The relations between brute facts are understood as horizontal, the sign relation, by contrast, as vertical. Every brute fact can be the X term of a sign function; nevertheless not all of them are in fact used this way. This is what Barthes (1968, 41) refers to as utilitarian and functional signs; simply put, it means that every X term of a sign function can be studied as a brute fact. The sounds we produce and call language, for instance, can be subjected to the exact same analysis as is applicable to any other acoustic signal.

The distinction between brute and institutional facts corresponds to the distinction between the physical and the semiotic and somehow reminds us of the traditional attributes given to the expression and content level of the sign, the former being sensible, the latter intelligible (see e.g. Jakobson 1949). The problem semiotics has with physical reality is that objects need not only function as perceptible bearers of sign function but are also something to which the sign refers. That is why semiotic discussion on the nature of the sign and how to clearly separate it from a non-sign very often ends up in a complete rejection of physical reality, reference, or, more generally, realism (see for example Devitt – Sterelny 1999, 265-270); while the definition of sign also refuses any kind of reference to brute facts or physical reality.

I am nevertheless of the opinion that there is no need for such rejection if we see reality as that which Peirce (see CP 4.536) refers to as the Dynamical Object, that is, as a fact that motivates sign function without necessity (or even chance) for being a part of it. An object can motivate us to give it a name a tree, for instance, but this does not automatically imply that such object is itself a sign, nor does it mean that such object completely lacks any relation to this sign-name. On the other hand, as Eco (1999, 65) points out, the Dynamical Object can serve as an expression term (or concrete token) of a sign function, so that we can perceive this object as the X term that stands for, for instance, an instance of the beauty of nature or, on the contrary, the expansion of human civilization. We can lie (or be mistaken) about Dynamical Objects, but we can use Dynamical Objects to lie only when they are treated as a sign token.

Physical reality consists of potentially infinite number of Dynamical Objects serving as an a priori for every possible sign function under a sole condition: that such an object becomes part of a sign function if and only if it is recognized as such (see Searle 1983, 163).

2   Representation and intentionality

Whereas the twitching boy has turned out to be the concern of neurology (or natural science), and that is where we shall leave him, his winking friend is of some further interest, namely his conspiratorial intent, which should attract our semiotic curiosity. The twitching boy twitches and cannot do otherwise, that is, he cannot lie with his eyelids. The winker, on the contrary, intends to inform his ally of, for instance, the right time to perform an arranged action (and he can lie about that, i.e. deliberately wink at the wrong time).

Throughout the whole of Searle's work there is a consistent claim that nothing can be perceived as a sign and at the same time unrelated to intentionality. In this point Searle's sign model departs from that proposed by de Saussure. De Saussure's complex of signifiant and signifié is on one hand defined in psychological terms but simultaneously rejects intentionality or "psychological truth" of communicative or speech acts. If we understand (see Searle 2002, 77) intentionality as a certain feature of our mental states that represents something other than itself, intentional object, or, as Peirce (see CP 4.536) would say, Immediate Object, than we also get to the heart of the nature of sign definition. The XYC relation is inconceivable without the notion of representation and the most striking of representative institutions is language.

I have said that every brute fact can serve as the X term in a sign function. There is however an intuition suggesting that some physical objects are more suitable to function in this particular position than others. This intuition has already shown up in the first sense of Dynamical Object, that of giving a name to an object. As Searle (1995, 60) claims, social reality is founded by the existence of language as a sign system. The necessary question is then, why? Within agentive functions there are these special cases Searle (1995, 21) called "representative" and as the most obvious example of such cases he gives language, the nature of which nature is to assign functions to sounds and marks. The central position of language and linguistic signs follows exactly from these formulations.

Representation is a synonym for intentionality and its manifestation is most obvious in language. Mostly, we do not use language without some intention and if we do, it should be of concern of specialized natural scientist. Likewise, the expression of linguistic sign shows some

peculiarity. First of all, the repertoire of linguistic expressions is limited (every language, for instance, has a closed phonological system). As language users we are truly homo economicus; we want to achieve as much as possible (preferably everything) with minimal exertion. In the process of searching for the best solution it naturally has to occur to the seeker that the most convenient of all the physical means in the world must be such as every human being can access readily at all times, that is, something that can be produced with the help of our own bodies.

Apart from this need for accessibility, there is yet another effort that seeks to identify a set of elements as limited as possible. It has something to do with the nature of our long-term memory (see Jackendoff 2003, 152) because it is much easier to remember a few phonemes and a few thousands of words than billions of sentences. The linguistic expression is a specially formed X term that satisfies both of these economies. When I say specially formed, I suggest that the X term of sign function need not to be taken from physical reality untouched. In many cases (maybe all of them) assigning a function is accompanied with some sort of intentional creativity on the side of X term. Semiotics often concerns itself with the economical nature of the expression plane while striving to find a similar principle on the content plane. At the same time, however, it tries to forget the variety of actual manifestations; when focusing on the problem with solely economy in mind, it misses the point.

Semiotics often answers its questions by using the smallest functional units, focusing on the compositional nature of signs (which parts constitute a certain whole?) and does not raise questions as to what we can do with it. Semiotics has a strong tendency to propose formally elegant descriptions and models, in an attempt to reduce the whole complex of semiosis to syntactic and semantic rules. From this point of view, speech act is not a semiotic term at all, if we understand semiotics as a closed discipline concerned with closed and immanent systems (see Ricoeur 1968, 120). This helps to explain, for example, why Searle's (1979, 1-29) taxonomy of illocutionary acts employs several criteria from which none can be perfectly matched to a respective semiotic economical/functional compositionality. Since general semiotics involves both closed systems (words) and open system (sentences, texts), the question of representation arises regardless of whether our language is economical or not and to reduce the sign problem to its economy means to give up the notion of sign as a function.

However, just as with language, every sign function must be intentional yet the intention in order to fulfill the wanting (of, for instance, the winker) has to be of a specific kind. If we get back briefly to the Dynamical Object a priori we can generalize the thesis by saying that there must be something prior to assigning functions, that is, we cannot create a sign function from scratch but we have to have something (be it brute fact or an already established sign function) upon which we set up the new one (the Y term of one function can be the X term of another etc., see CP 1.339 and Searle 1995, 82-83). From this emerges the famous distinction between semiotics of signification and semiotics of communication (see Eco 1979) which is analogous to the thesis that "representing intentions are prior to communication intentions" (Searle 1983, 166).

If we agree with the relative autonomy of representation and communication we can subsequently assert that while representation can be individual, the communication is necessarily collective and social. In principle then, we can distinguish between those sign functions which were created with the intention of including only the creator (I-intentional signs) and those which were created with some sort of collective intention in mind (We-intentional signs). We can call both of them signs because they meet the semiotic requirements mentioned so far (XYC relation and intentionality).

It seems to me that this division could be plausible mainly in the restricted domains of certain human activities (such wherein we are interested in someone's creative, i.e. I-intentional sign-making act) but in the end we have to admit that the separation of representation and communication is rather virtual and it is so due to the language-centered nature of social reality or, to put it more generally, we-intentionality necessarily comprises signification as well as communication; there is no "we intend" without communication which presupposes signification. Great example of this mechanism would be a scrabble-like game where the player's goal was to create a sign function acceptable collectively (or at least by his opponent) from already existing things. But is this not the game we all play all the time?

If Searle is right (and I think he is), every person can, individually, impose a function arbitrarily upon whichever object they desire. This is however not a sign. I can wink all day long while I-intending my winking as standing for whatever I wish, it can even represent different "meanings" in different contexts and, as a result of this, formally

satisfy the sign model definition. Still, it is not a sign. To use a more appropriate example, should an individual means of mine assigns the function "The President of the Czech Republic" to my father, it would completely lacks the collective dimension; it would not constitute an (institutional) fact.

The second semiotic condition (see Searle 2002, 102) is as follows: means can be individual but ends have to be collective. The X term becomes a part of sign function if and only if we do not extrapolate I-intentions to we-intentions (see Searle 2002, 93) and consider we-intentions as descriptive primitives of every sign function as its inseparable part.

3   Arbitrariness and constitutive rules

The winker wants to tell his ally that just now is the right time to perform certain action. But how does his ally know that he should interpret the wink like this? We have to admit that a sign, apart from being a representation, is also a constitutive rule.

In semiotics, there is a strong tendency to neglect this feature. Constitutive rules are, in a sense, opposed to de Saussure's notion of arbitrariness which states that there is no motivation that causes X to count as Y in C. This notion led to a natural critical reaction (and in some cases rejection of the whole notion of arbitrariness) because what we really want to know is not that X does not require certain Y (in both we- and I- sense) but rather why it is the case that we count X as Y in C.

Actually, it seems that arbitrariness tries to resolve the puzzling question of how the signs relate to the real world, e.g. when de Saussure states that there is no natural connection between expression and meaning. Searle (and some critics of arbitrariness), on the contrary, discusses this problem as a problem of rule-governed connection on the level of institutional facts. Such type of facts are conditioned by we-intentionality that brute facts lack, arbitrariness in de Saussure's sense, therefore, is not an analytic term, it is a fact itself that must be analyzed with special attention to the consequences for we-intentionality.

De Saussure (1959, 71) probably sees these consequences when he states that "[t]he signifier, though to all appearances freely chosen with respect to the idea that it represents, is fixed, not free, with respect to the linguistic community that uses it". De Saussure however (1959, 71) understands this relation as "a thing that is tolerated and not a rule to

which all freely consent". It would seem that any sign function we have is simply there and we have to accept it without any possibility to influence it (again, in we- and I- sense).

When de Saussure defines arbitrariness, it seems that what he has in mind is the sign function. What he speaks about, however, is the principle of horizontal articulation, i.e. the relations between Y terms or XY compounds. The horizontality completely sidelines the sign function and tries to proceed with the same methodology the natural sciences use on the level of brute facts. De Saussure, though, speaks about the sign employing psychological terms while the core of his semiological project lies in an attempt to establish basic logical rules of horizontal articulation. The semiological project, therefore, is not grounded by X counts as Y in C, but rather that Y1 differs from Y2, Y3 … YN (or XY1 differs from XY2, XY3 … XYN; it is not at all clear), and this principle of differentiation is at the same time the founding principle of sign.

De Saussure overshadowed the vertical sign function in favor of relations between individual separate sign constituents. The "counts as" is transformed to the "differs from" and even though this systematicity must no be neglected (see Searle 1995, 35-36) it should not be overestimated, either, especially when describing the constitutional sine qua non of sign functions.

The constitutive relationship was put forward by Peirce with his notion of interpretant (in the mediating sense described in CP 1.553). This Peirce's term expresses not only the representative nature of signs in its clearest form (see CP 1.555); it therefore also allows the constitutive rule to be explained using this term.[1] Interpretant is a general guarantee mechanism of sign function stability best seen in or when using language. Every institutional fact except for linguistic signs results from the so-called status function declarations (see Searle 1995, 34), i.e. a special type of speech act that creates signs by its successful perfor-

_____

[1]   Of course, the huge amount of commentary on Peirce's semiotics I say nothing of here is characterized by endless application of Peirce's terms to nearly everything which often results in complete indistinctiveness. But if we nod in deep understanding to Whitehead's "philosophy as a series of footnotes to Plato" we should also nod (at least methodologically) to Shalizi's (1998) "American thought is a series of footnotes to Peirce".

mance. The reason behind this is that every possible sign can possess interpretant(s) made up solely of linguistic signs.[2]

Some confusion could arise between signs the interpretants of which do have constitutive function and are formulated in language, and descriptions of physical objects which can also employ language but in this case there is no interpretant in the constitutive sense. "[S]cience is a linguistic representation of experience", as Jakobson (1971, 690) puts it, and, therefore, science is itself an institution. The problem is that literally every brute fact can serve as the X term of a sign function but, on the other hand, sometimes a brute fact is just a brute fact. The paragon of brute fact description suggests its Dynamical-Object nature. Brute facts completely lack interpretant, which is not the case with descriptions (in the form of judgments, diagrams etc.). This is obvious in closer look at the linguistic sign; its function is a model of meaning but not the act itself we perform using this sign function (see Searle 2010, 14). Utterances are not SF declaration's aftermath but their very existence is based on meaning, or simply language. The language is already a language and that is why it does not require any previously existing one, as Searle points out (see Searle 1995, 72).

To conclude, we can say that interpretant is a mechanism that plays its role both in social institutional and social non-institutional facts. Sometimes it can happen that a thing is created without intention or effort to assign any function to it. As time goes by, however, the society (or social reality) turns such creation into the X term of a sign function regardless of creator's indifference towards (or even explicit opposition to) assigning such a function to it. The same is possible for the reverse. This should not be seen as an obstacle but rather an inherent feature of the social reality's ontology. "The object", Searle (1995, 36) points out, "is just continuous possibility of the activity".

The third semiotic principle directly emerging from Searle's writings is the constitutive and normative nature of signs without a necessity to be static and forever unchangeable (for some discussion see e.g. Koťátko 1998). Of course, sign description is easiest in the domain where normative consequences of signs are evident. When Guiraud (1978, 13) claims that "[t]he greater the redundancy, the more the com-

_____

[2]   If all signs are in some sense derived from or dependent on language, it is understandable why there are attempts to describe linguistic sign system by means of horizontal logic.

munication is significant, closed, socialized and codified; the lower the redundancy, the greater the information and the more open, individualized and decodified the communication", it indicates the tension between the I-intentional signs which show less normative power (if any) and the we-intentional signs where the consequences are inherent in their communicative nature.

4   Conclusion

What has been said above is only a brief sketch that interweaves some of Searle's theses with those found in the traditional semiotics. The reason why Searle's theory should matter to semiotics lies mainly in that it offers general concept for a description of (social) reality but at the same time comprises a theory of language. The theory can be applied with success to words or sentences as well as to other institutional facts (see, for example, Searle 2010, 91-92). Moreover, Searle offers a great starting point in the form of the relational triad which is well established in semiotics.

It seems to me that everything Searle is saying has a common denominator, sign, which, therefore, has three characteristics. (1) It is an institution and cannot be reduced to physical objects or laws. (2) It is a representation, that is, sign requires human agent capable of assigning functions upon objects. (3) It is a constitutive rule, an interpretant that homologizes use and recognition of objects as signs.

Naturally, there are other, different kinds of signs (I completely disregard, for instance, questions of icons or indices) but the related problems are connected to more specific semiotic investigations and do not have general solutions (apart from generic terms such as similarity which themselves need further analysis). What is important here is that if we consider something as a sign we can, at the same time, say that it is a fact. Searle's semiotics (see Searle 1995, 7-9) draws a clear line between subjective and objective in both the epistemic and the ontological sense.

Consequently, there are many ways in which signs can be described within this frame but the most important of them is that they are facts. In this sense, I appreciate Fish's response to Alan Sokal where Fish states the sign-fact thesis in the form of little catechism (as he calls it):

Are there balls and strikes in the world? Yes. Are there balls and strikes in nature (if by nature you understand physical reality independent of human actors)? No. Are balls and strikes socially constructed? Yes. Are balls and strikes real? Yes. Do some people get $3.5 million either for producing balls and strikes or for preventing their production? Yes. (Fish 1996, A23)

Fish points out that when something is a sign it does not mean that it is not real. Similarly, Searle's theory is focused on the ontology of signs and it seems to me it is probably one of the most interesting pieces of the history of semiotics that does not use semiotic terminology. If we admit that philosophy of language is a branch of philosophy of mind (as suggested in Searle 1983, 160), it could strengthen our belief that there is a common ground for semiotics and philosophy.

Department of General Linguistics
Faculty of Arts
Palacký University
Křížkovského 10 / Křížkovského 14
Olomouc 771 80
Czech Republic
v.gvozdiak@gmail.com

References

BARTHES, R. (1968): *Elements of Semiology*. New York: Hill and Wang.

COBLEY, P. (ed.) (2010): *The Routledge Companion to Semiotics*. London – New York: Routledge.

DEVITT, M. – STERELNY, K. (1999): *Language and Reality*. Oxford: Blackwell Publishers.

ECO, U. (1979): *A Theory of Semiotics*. Bloomington: Indiana University Press.

ECO, U. (1999): *Kant and the Platypus. Essays on Language and Cognition*. San Diego – New York – London: Hartcourt.

FISH, S. (1996): Professor Sokal's Bad Joke. In: *The New York Times*, May 21, A23.

GUIRAUD, P. (1978): *Semiology*. London – Henley – Boston: Routledge & Kegan Paul.

JACKENDOFF, R. (2003): *Foundations of Language. Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.

JAKOBSON, R. (1949): The Phonemic and Grammatical Aspects of Language in Their Interrelations. In: Lejeune, M. (ed.): *Actes du Sixième Congrès International des Linguistes*. Paris: Klincksieck, 5-18.

JAKOBSON, R. (1971): Linguistics in Relation to Other Sciences. In: *Selected Writings II. Word and Language*. The Hague – Paris: Mouton, 655-696.

KOŤÁTKO, P. (1998): *Význam a komunikace*. Praha: Filosofia.

MORRIS, Ch. W. (1938): *Foundations of the Theory of Signs*. Chicago: The University of Chicago Press.

PEIRCE, Ch. S. (1931 – 1935, 1958): *Collected Papers of Charles Sanders Peirce*. Cambridge (Mass.): Harvard University Press.

RYLE, G. (2009): The thinking of thoughts: What is 'le Penseur' doing? In: Ryle, G.: *Collected Papers. Volume 2. Collected Essays 1929 – 1968*. London – New York: Routledge, 494-510.

RICOEUR, P. (1968): Structure – Word – Event. *Poetics Today* 12, 114-129.

DE SAUSSURE, F. (1959): *Course in General Linguistics*. New York: Philosophical Library.

SEARLE, J. R. (1969): *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

SEARLE, J. R. (1972): Chomsky's revolution in Linguistics. *The New York Review of Books* 18, 16-24.

SEARLE, J. R. (1979): *Expression and Meaning*. Cambridge: Cambridge University Press.

SEARLE, J. R. (1983): *Intentionality*. Cambridge: Cambridge University Press.

SEARLE, J. R. (1995): *The Construction of Social Reality*. New York: The Free Press.

SEARLE, J. R. (2002): *Consciousness and Language*. Cambridge: Cambridge University Press.

SEARLE, J. R. (2010): *Making the Social World. The Structure of Human Civilization*. Oxford: Oxford University Press.

SEBEOK, T. A. (ed.) (1986): *Encyclopedic Dictionary of Semiotics*. Berlin – New York – Amsterdam: Mouton de Gruyter.

SHALIZI, C. R. (1998): *Review: Error and the Growth of Experimental Knowledge by Deborah G. Mayo*. Available:

http://www.phil.vt.edu/dmayo/personal_website/Shalizi_review_EGEK.pdf.

How to Make the Concepts Clear: Searle's Discussion with Derrida

Tomáš Koblížek

The Academy of Sciences of the Czech Republic, Prague

Abstract: The first part of the paper deals with the key question of the Searle-Derrida debate, namely, with the question of conceptual "exactness" and applicability of concepts to facts. I argue that Derrida makes a strict distinction between the exactness in the realm of concepts and the exactness in the realm of facts. Supposing that it is not correct to argue against him – as Searle does – that concepts cannot be exact because there are no strict boundaries between facts. The second part of the paper deals with a distinction used by John Searle: The distinction between linguistic meaning and speaker's meaning. According to Searle linguistic meaning is constituted outside a particular context of use whereas speaker's meaning is embedded in a particular situation. I argue this distinction is problematic as far as any meaning is constituted in a particular utterance and in a particular context of use.

Keywords: Background, concept, speaker's meaning, linguistic meaning, utterance.

1

The famous debate between John Searle and Jacques Derrida came to an end nearly twenty years ago. John Searle's last reply appeared in his *Construction of Social Reality* (1995), Derrida's final contribution can be found in his "Afterword" to *Limited Inc* and is seven years older (1988). The polemics dealt mainly with principles of the speech act theory, the "iterability" of signs, and the complexity of speaker's intention and I will not examine it in its entirety. I would only like to return to two questions discussed in *Limited Inc* and in two texts by John Sear-

le, namely in his review of Jonathan Culler's *On Deconstruction* (1983) and in his (1994) essay "Literary Theory and its Discontents". The first question concerns the boundaries between concepts, the other question deals with the meaning of utterance. As we shall see, both questions are closely related since the answer given to the first can function as a prelude to answering the second.

As far as the question of concepts is concerned, John Searle addresses to Derrida the following objection:

> First there is the assumption that unless a distinction can be made rigorous and precise it isn't really a distinction at all. Many literary theorists fail to see, for example, that it is not an objection to a theory of fiction that it does not sharply divide fiction from nonfiction, or an objection to a theory of metaphor that it does not sharply divide the metaphorical from the nonmetaphorical. On the contrary, it is a condition of the adequacy of a precise theory of an indeterminate phenomenon that it should precisely characterize that phenomenon as indeterminate; and a distinction is no less a distinction for allowing for a family of related, marginal, diverging cases. People who try to hold the assumption that genuine distinctions must be made rigid are ripe for Derrida's attempt to undermine all such distinctions. (Searle 1983, 78)

Later John Searle explicitly applies these claims to distinctions between concepts as he points out that "most concepts and distinctions are rough at edges and do not have sharp boundaries", or that it is "generally accepted that many, perhaps most, concepts do not have sharp boundaries, and since 1953 we have begun to develop theories to explain why they *cannot*" (Searle 1994, 637, 638). According to Searle, the looseness of boundaries refers to the fact that there are "marginal" or "diverging" cases which complicate any clear conceptual distinction. Perhaps Derrida and his followers merely neglect this fact and take the opposite view regarding the concepts as something clear and distinct – thereby they commit a fundamental mistake.

This objection is especially disconcerting for Derrida and he expends much effort to refute the argument. Let's leave aside his claim that philosophers have always held that in the order of concepts "when a distinction cannot be rigorous or precise, it is not a distinction at all" (Derrida 1988, 123). Reference to historical background is surely not the main point of Derrida's defense even though the idea of tradition

and its power generally plays a key role in deconstruction. We must rather focus on his assertion concerning the applicability of concepts to facts. Derrida explicitly claims that empirical difficulties "do not, in fact, exclude the possibility of a juridical-theoretical process leading to an essential definition," and he provides the following example: "if one wishes to know what conditions are necessary for a promise, for instance to be a promise, it ultimately matters little whether or not *in fact* a promise has ever existed, or whether one has ever been *actually* discovered which would fully and rigorously satisfy the requisite conditions" (Derrida 1988, 69; emphasis mine).

Here we find the crucial distinction referred to in Derrida's demand of "rigorous boundaries" between concepts. The "crystal clarity" of concepts apparently concerns the exactness of their *definition*, not their *application* to facts. According to Derrida, there are on one hand concepts as specific products of idealization – concepts determined by their definition or their definitional features – and on the other hand there are facts which satisfy the conditions set up by the definition only to a degree, "more or less". This is to say that in principle we can clearly define what it means "to promise", "to declare", or "to lie", yet, in each case of a particular use of the concepts, as we are obliged to apply the concept to specific utterances, there will always be some marginal phenomena or undecidable facts which will fail to perfectly fulfill the conditions of the concept. We know what promise is *per definitionem*, yet we need not to be sure – and usually we are not – if this or that utterance is *de facto* a promise or rather something else.

It is important to notice that this fundamental distinction is not affected by the mutability of definitions. The fact that we can change definitions has nothing to do with the simple truth that we can *simultaneously* think (1) the rigorous distinction between concepts and (2) the undecidability of their application. Derrida returns to this point in several passages of *Limited Inc* and he always speaks out decidedly. For example, in the "Afterword" he once again points out: "Every concept that lays claim to any rigor whatsoever implies the alternative of 'all or nothing'. *Even if in 'reality' or in experience everyone believes he knows that there is never 'all or nothing'*, a concept determines itself only according to 'all or nothing'" (Derrida 1988, 116; emphasis mine) – "all or nothing" indicating: either there are strict boundaries or there are no boundaries at all.

When John Searle returns to these claims he seems to overlook the distinction between (1) conceptual boundaries and (2) the clarity of particular cases to which we apply the concepts. (As we have seen, the concepts are supposed to be strict whereas the particular cases might lack the requested clarity.) To give an example of John Searle's reading of the argument it will suffice to look at his essay on "Discontents". Here he rebuts the "all or nothing" approach to concepts by pointing at their applicability. That means, the boundaries between *concepts* are shown to be loose by reference to different *facts* to which we apply the concepts. Right in the opening passage of the essay we find a statement confusing both levels of analysis: "I pointed out that it is not necessarily an objection to conceptual analysis, or to a distinction, that there are no rigorous or precise boundaries to the *concept analyzed* or the distinction being drawn. It is not necessarily an objection even to theoretical concepts that they admit of *application* more or less" (Searle 1994, 637, emphasis mine). In a similar way, few paragraphs later, Searle refuses the idea of purity of *concepts* by pointing out the "un-purity" of particular *cases*. He writes: "It is clear from this discussion that Derrida has a conception of '*concepts*' according to which they have crystalline purity that would exclude all the marginal *cases*" (Searle 1994, 637; emphasis mine). Derrida indeed possesses the notion of "crystalline pure" concepts, however, these concepts – when applied to facts – actually *admit* of marginal cases: A marginal case does not imply unclarity of a concept. Moreover, it is clear from the "Afterword" that Derrida is aware of the empirical argument against conceptual sharpness but he is determined not to use it: "To this oppositional logic [that is, to the oppositional logic of concepts], which is necessarily, legitimately, a logic of 'all or nothing', and without which the distinction and the limits of a concept would have no chance, I oppose nothing, least of all a logic of approximation, a simple empiricism of difference in degree" (Derrida 1988, 117). Evidently, for Derrida, the counterargument based on undecidable empirical facts is of no use. It violates the distinction between conceptual and empirical (factual) clearness.

2

All this is not to say that the conception of two separated "worlds", "the world of concepts" and "the world of facts", is unproblematic and resistant to doubts. On the contrary, such a conception must be

re-considered if we are aware of some basic problems in philosophy of language and general linguistics. In this part I will argue that there is no completely impervious boundary between concepts and facts, or "factual context", insofar as concepts are available to us only by means of language. I believe that the "linguistic nature" of concepts breaks the boundary between the two realms. In this context, we can make use of the Searlian definition of utterance. However, to prove the claim, we have to suspend the difference between the speaker's meaning and the sentence meaning which John Searle uses in his semantics.

As a starting point we can use Searle's example which illustrates his notion of "Background". Searle writes:

> Consider, for example, the utterance, "Cut the grass". Notice that we understand the occurrence of the word "cut" quite differently from the way we understand the occurrence of "cut" in "Cut the cake" (or "Cut the cloth", "Cut the skin", and so on) even though the word "cut" appears univocally in both sentences. This point is illustrated if you consider that if I say to somebody, "Cut the cake", and he runs a lawnmower over it, or if I say, "Cut the grass", and he runs out and stabs it with a knife, we will, in each case, say that he did not do what he was literally told to do. How do we know, as we do know, which is the correct interpretation? We do not have different definitions of the word "cut", corresponding to these two occurrences. We understand these utterances correctly, because each utterance presupposes a whole cultural and biological Background (in addition to a Network of beliefs, and so on). (Searle 1994, 640)

In order to re-consider the role of concepts we must focus on the assertion that we do not possess "different definitions of the word 'cut', corresponding to these two occurrences." Such a claim is perfectly correct: There is no special definition of the verb "cut" for the case of cutting cake, nor is there a specific definition of the verb "cut" for the case of cutting grass. As John Searle puts it, we rather interpret the utterances using the Background: the utterance leans on "a set of background capacities, abilities, presuppositions, and general know-how" which enable us to understand (cf. Searle 1994, 640).

Nevertheless, Searle's approach is quite peculiar at this point. He claims that there are no corresponding definitions, yet he admits that there is a common level of meaning which is essentially different from the particular "speaker's meaning, as determined by the speaker's intentions on particular historical occasions" (Searle 1994, 647). He ex-

plicitly talks of "the sentence" or "the word meaning" and identifies it with linguistic convention. In this vein, later in the discussion, he points out that "the meaning of a text *can be examined quite apart* from any authorial intentions, because the meaning of the text consists in the meanings of the words of which it consists" (Searle 1994, 652; emphasis mine). The meaning indicated here is not the utterance meaning of the speaker but, perhaps, the meaning we find in dictionaries.

Nevertheless, we could ask how such a type of meaning is available to us. How can we "examine" the meaning which the word "cut" possesses outside a particular utterance, that is, outside an utterance pronounced in a particular context or a historical situation? I would like to put forward the following suggestion: Even the linguistic definition of the word "cut" requires the word to be uttered, situated in "particular historical occasions", and, therefore, the word or the sentence meaning does not represent a type of meaning which would be essentially different from utterance meaning. Or, to put it the other way round, in order to save the idea of word or linguistic meaning one would have to put forward such a definition of the word "cut" which would be absolutely detached from any "particular historical occasions". One would have to present a meaning that would be fully comprehensible and yet this comprehension would not be supported by any particular context of use. It seems clear that such an effort must be in vain and that such a definition is impossible. The so-called linguistic meaning or conventional sentence meaning is not independent of the situation of speaking. Rather, it is the meaning of an utterance *which we often use*: *the typical utterance meaning*.

To make the assertion clearer let's pick up another example discussed by John Searle in his polemics with Jacques Derrida. This time the discussion deals with the meaning of Nietzsche's note "I have forgotten my umbrella" that can be found in *Nachlass* of the author. Here John Searle denies Derrida's claim that the sentence might have no meaning, or as Derrida writes: "Because it is structurally liberated from any living meaning, [*vouloir-dire vivant*], it is always possible that it means nothing at all or that it has no decidable meaning" (quoted from Searle 1994, 661). Searle's main counterargument is based in the very notion of "conventional meaning", or in the notion of "sentence type" which is identified with the linguistic meaning of a sentence. The objection runs as follows: "The German sentence type has a conventional meaning in German. Given the Network and the Background, the inter-

pretation of sentence meaning is quite determinate. In a different Background culture, where all umbrellas were made of chocolate and eaten for desert after use in rainstorms, the literal sentence meaning could be understood differently (it might mean: I have forgotten the taste of my umbrella); but given the existing cultural, biological, and linguistic situation in the late nineteenth century, the literal interpretations are unproblematic" (Searle 1994, 661). In a similar way, Searle goes on to differentiate between "speaker's meaning" and "sentence meaning". He writes: "From the fact that Nietzsche might not have meant anything by the production of the token (speaker meaning) it does not follow that the token might 'mean nothing at all' (sentence meaning)" (Searle 1994, 661 – 662). Searle's counterargument is clear: As far as there is a type of meaning (linguistic meaning or sentence meaning) independent of the meaning intended by the author, the sentence will always mean something – no matter whether or not Nietzsche had anything in mind. Such a meaning – type meaning or linguistic meaning – is determined by *conventional* network and background presuppositions.

Is there anything wrong with the claim? At this point, we must carefully judge the particular case. It is quite correct to assert – against Derrida – that Nietzsche might not have meant anything and that the sentence is still meaningful. Yet, we should be aware of the fact that Nietzsche's note is not meaningful simply because it is in German, that is, because it has a linguistic meaning which is essentially different from the particular meaning of the utterance. In fact, when John Searle points to the Background and the Network which determine the interpretation of the conventional meaning, he is pointing to a *particular context* as determining the meaning of the *utterance* – not determining a different type of meaning, the linguistic meaning. Nonetheless, as in the case of the speaker's meaning, he must think of a particular situation or context of use in order to obtain the so-called linguistic meaning. At the very least he must imagine somebody intending the meaning in such a context. Perhaps, the context would be: it is raining and the umbrella is an instrument I can use in order to stay dry.

Why is it, then, that John Searle would still like to talk of *two types* of meaning? Does he simply deny the unity where no difference can be found? It seems his approach is based in a specific procedure: John Searle usually analyzes an utterance in *two* different contexts and by this analysis he obtains two meanings. However, retrospectively, he claims that the two meanings are co-present in *a single* context of use

and that they represent *two types* of meaning. To be more specific: At first, the utterance is set up against a more typical context and a more typical network of beliefs. By this procedure we obtain the so-called linguistic or conventional meaning. In the "Discontent" essay the utterance "the window is open" serves as an example. The linguistic meaning is identified by reference to the common context of use: there is an open window (see Searle 1995, 645). Subsequently, the same utterance is confronted with a less typical or an "individual" occasion. By this procedure we obtain the so-called speaker's meaning. Here John Searle provides an example of diplomatic context where "the window is open" could mean "there are opportunities for further negotiations" (Searle 1995, 646). Yet, it appears that the difference between these two meanings – linguistic meaning and speaker's meaning – is not a matter of *type* but rather of *typicality*: it is a difference *in the degree of typicality* between two utterance meanings. In one particular context, the utterance refers to what it commonly refers, i.e. to window as an object which we open to get fresh air. The other context is specific: it concerns the situation of a diplomatic meeting. However, John Searle would still insist that the linguistic meaning – the common meaning – can be found in the diplomatic utterance too, no matter what the diplomats think. I would rather say: the so-called linguistic meaning cannot be found in this linguistic unit automatically, it cannot be found there as some kind of a permanent semantic layer. The utterance would have to be projected against another (common) context where the so-called linguistic meaning is constituted.

Why spend so much time dissolving the difference between the two types of meaning? Why should such a thing matter in philosophy of language? There are various reasons but especially one in particular could be put forward for discussion. It is worth noticing that the dissolution enables us to eliminate the problem of how we should understand the *relationship* between the two types of meaning in an utterance. John Searle often refers to this relation as something complex or complicated but he gives no explanation of how these two types of meaning can exist together in one linguistic unit (cf. e.g., Searle 1994, 647 or 659). The analysis becomes much easier if we do not operate with two types of meaning. It suffices to make clear that we either interpret the meaning of utterance as referring to typical occasions, or else we deal with utterances in a situation which is quite unusual. From this it follows that linguistic and speaker's meaning are not two semantic lay-

ers enveloping one other in a single utterance. There is only one type of meaning, sometimes more and sometimes less common or typical.

Institute of Philosophy
The Academy of Sciences of the Czech Republic, v. v. i.
Jilská 1
110 00 Praha 1
Czech Republic
tomas.koblizek@gmail.com

References

DERRIDA, J. (1988): *Limited Inc*. Transl. S. Weber and J. Mehlham. Northwestern UP: Evenston.

SEARLE, J. R. (1994): Literary Theory and its Discontents. *New Literary History* 25/3, 637-667.

SEARLE, J. R. (1983): The Word Turned Upside Down. *The New York Review of Books*, 74-79.

IV. Fictional discourse, metaphors

Searle's Approach to Fiction
(Extending the Concept to Other Media)

Jiří Koten

Jan Evangelista Purkyně University, Ústí nad Labem

Abstract: The essay summarizes crucial propositions of John Searle's approach to fiction and extends the analysis to other genres, specifically to drama, photograph and film. For Searle, novelists pretend to make assertions, because they need to make use the effect inherent in this sort of speech acts – to represent a state of affairs. We believe that all fictions arise as imitations of authentic representation: a fictional photograph imitates a documentary photograph that is the image captured with the help of photographic film or digital media. A fictional film imitates real people and real events recorded on a camera. Fictional film characters only exist, because the film-makers pretend that they have documented them. Fictions are a part of the social universe: we treat fiction according to the rules and habits we have acquired as members of the society. Fictions are also capable of imitating the effect of authentic representations: novels and films achieve to provoke real emotions.

Keywords: speech act theory, John Searle, fiction, pretense theory, representation.

In this essay, I am going to reflect on John Searle's paper *The Logical Status of Fictional Discourse*, which was first published in the journal *New Literary History* in 1975 and later included in the book *Expression and Meaning* (1979). Searle's paper dealt with illocutionary characteristics of literary fiction. Today I will attempt to outline Searle's conclusions and to demonstrate that their relevance goes beyond the borders of literature and its system. I will also try to elaborate on some of Searle's partial conclusions.

John Searle views the use of language as performing speech acts whose successfulness is conditioned by complying with constitutive rules. For example, if we want to make a successful assertion, the sen-

tences we pronounce must be true, verifiable and sincere. In *The Logical Status of Fictional Discourse*, Searle compares a newspaper report taken from the New York Times with an excerpt from Iris Murdoch's novel. At first sight the style of the newspaper article seems to be completely the same as that of the novel's narration. However, while the newspaper report can be seen as a successful assertion since it observes its constitutive rules, the novel passage does not respect these same rules (the speaker does not guarantee the veracity of the fictional utterance; she does not represent the actual state of affairs, etc.). By comparing these two texts, Searle finds out that authors of fiction only *pretend*: "By pretending to refer to people and to recount events about them, the author creates fictional characters and events" (Searle 1979, 73). Searle's observation is penetrating: what lies at the core of a fictional utterance is not the aim to deceive the recipient by a false assertion, but to make use of the effect inherent in statements – to represent a state of affairs. Although literature *cannot imitate reality* directly, it *can perfectly imitate an utterance about reality*. Therefore, fictional utterance employs the form of assertion even though it actually fails as an assertion. In this way, authors of fictions achieve a special effect: their readers are aware of the fictitiousness of this speech act, but – in accordance with the rules of this type of language game – they read the fictional account *as if* it was an assertion. During the act of reading, readers imagine in their minds, at least temporarily, the circumstances to which the novel refers as if it was the real world, even though they know that the reference world is not the real world.

Since the publishing of *The Logical Status of Fictional Discourse*, Searle's argument has appeared in a similar form in a great number of significant works dealing with questions of literary fiction. Richard Ohmann in (1971) regards the illocutionary force of fictional discourse as mimetic; Barbara Herrnstein-Smith in (1978) concludes that genres of literary fiction imitate genres of factual writing (authors of novels pretend that they are creating autobiographies, memoirs, etc.). Furthermore, the conception of fictional discourse as imitation of an authentic speech act has been reflected in semiotic theories: for example, Felix Martínez-Bonati portrays discourse of literary fiction as an iconic depiction of sentences (or as a "pseudo-sentences"; see Martínez-Bonati 1981, 78-79) which is in contrast with the linguistic denotation of objects in the empirical world. The fictitiousness of literary language is char-

acterized by the fact that it does not refer to reality but only depicts a communication act.

Searle's theory of fiction introduced in *The Logical Status of Fictional Discourse* is, in my opinion, valid for almost all kinds of fiction and not just for narrative or literary fictions. Searle himself considers the case of dramatic fiction, where, according to him, the actors take part in the pretending: "Here (in the case of drama) it is not so much the author who is doing the pretending but the characters in the actual performance" (Searle 1979, 69). I personally think that dramatic fiction is not as different from narrative fiction as it might seem. I think that the difference between narrative and dramatic fiction lies entirely in the manner of authorial pretending. While narrative fiction is based on a sequence of feigned assertions that we assign to the central narrator within fiction, there is no narrator in drama. Instead, the author creates monologues and dialogues of characters, which, however, have absolutely the same status as the narrator's assertions: they imitate authentic speech acts, such as assertions, warnings, promises, requests, orders, verbal expressions of states of mind and emotions, etc. The characters' utterances (and this is where drama differs from a narrative) are complemented by a secondary text, which is not intended for the (readerly) audience, but for the actors. Searle notices that unlike the characters' utterances, these stage directions are seriously meant speech acts. They are instructions that oblige the actors to act in a certain way on the stage. Yet the text of the stage directions does not disturb the readers even if they only *read* the play: through the directions, they are informed about what each character is doing. To come back to the original proposition by Searle, that is that dramatic fiction is not based on the authors' pretending but on the pretending of the actors, I think that this original argumentation merges the creation of dramatic fiction with its performance on the stage. From my perspective, dramatic fiction is produced in the same way as narrative fiction: it is based on a creative authorial imitation of speech acts.

I believe it is beyond all doubt that fictions arise as imitations of authentic representation. According to Searle, it is non-deceptive pretending, and in the case of verbal fictions it is based on the illocutionary stance of the author. The fact that fictitiousness cannot be recognized based on the nature of the depiction itself but that it is dependent on the authorial intention which the recipient must be able to identify applies, in my opinion, to other media as well. For instance, photography is a

means of capturing images with the help of photographic film or digital media. Everybody who has ever taken a holiday photo knows that a photograph can preserve an authentic image of an object. A photo of Sean Connery in a tuxedo and with a small Beretta pistol is an authentic image as well. However, if the caption says that the image does not represent a person called Connery, but a person called James Bond, I can assume that this is fiction because an instance, analogical to the author of literary fiction, has arranged the picture. The fiction is created by the person who pretends that the photograph is an authentic representation of Bond. This happens in the same way in film. Film can record moving pictures and sound in a documentary manner. Incidentally, the first films depicted mundane scenes such as the arrival of a train at the station, as in the film projected by the Lumière brothers in Paris in 1896. In keeping with Searle's proposition, cinematic fiction could also be described as pretending: the creators of the film pretend that through the medium of film, they have recorded real events as they occurred. The fictitiousness of the events on screen therefore stems from the fact that they imitate the version of reality captured by the camera; in other words, film characters and events only exist because the film-makers pretend that they have documented them.

Literary theory is often puzzled by our understanding fictions, i.e. by the fact that we can identify them and read them in an appropriate manner. Gregory Currie and Kendall Walton invented the term *make-believe* to describe the attitude we assume towards fictions (see Currie 1990; or Walton 1990). Make-believe does not stand for believing in the veracity of an assertion but rather for the willingness to employ our imagination in dealing with the content that is communicated to us. In a manner following Searle's, Currie is attempting to determine precise rules enabling the recipient to identify a fictional utterance. In my opinion, analyses of this kind are somewhat complicated and also redundant, while the key significance lies in the fact that we can appreciate the nature of artistic fictions. Literature is a part of the social universe which John Searle describes in his books (see e.g. Searle 1997): we treat fiction in the form of literature and film according to the rules and habits we have acquired as members of the society. Novels and cinematic stories are collectively perceived not as representations intended to capture reality but rather as a special kind of pseudo-representations intended to invoke a certain type of an aesthetic experience (such as amusement, enlightenment or emotion). Literary and cinematic fictions

constitute a common part of social reality: we know in what way we are supposed to read or watch them in the same way as we know that we are supposed to use money to pay for goods or a park bench to rest.

With this remark about experiencing fictions I have arrived to the last point I would like to consider. Towards the end of *The Logical Status of Fictional Discourse*, Searle poses a question of why we should even rack our brains about fictions, that is, representations (be it verbal or other) that we know to be pretended. The philosopher concludes that fictions, despite their "non-seriousness", facilitate the telling of serious messages. Even though there is no doubting the correctness of this conclusion, I would like to consider another, far more mundane explanation. Quite simply, we enjoy reading fictions – they can bring us very real pleasure. While reading a fictional story, I enter a communicative situation that could be described as a communicative exchange with *double illocutionary aspect*. To stay with Searle's example: I am reading a novel in which Doyle pretends to be John Watson, who is recounting a true story. At the same time, however, I temporarily forget that I am dealing with literary fiction by Conan Doyle, and I read the novel as though it presented true memories of Dr. Watson, relating sincerely and truthfully the adventures he experienced with Sherlock Holmes. Stories such as *The Hound of the Baskervilles* provoke the feeling of suspense and horror, and yet this is a pleasant kind of fear. How is that possible? I believe that the answer lies in the specific character of fictional communication. Fiction is capable of imitating the effect of an authentic assertion: we are worried about fictional characters, we feel for their suffering, and occasionally they even make us cry. The fictional narrators' accounts portray the experience of somebody else and have – like speech acts – their own perlocutionary effects. However, the compassion, excitement and horror of fictions are provided in adequate doses: even though we can get carried away at times, eventually we will always realize that we are only dealing with a fictitious representation – that Watson, Holmes and the beast wandering around Baskerville only exist because Doyle has created them by a sequence of pretended assertions. Therefore, Doyle's pretended assertion has perlocutionary effects of its own: we can appreciate it as an accomplished work of art, as an impressively constructed fictional representation.

Department of Czech Studies
Faculty of Education
Jan Evangelista Purkyně University
***České** mládeže 8*
*400 96 **Ústí** nad Labem*
Czech Republic
jiri.koten@ujep.cz

References

SEARLE, J. R. (1979): The Logical Status of Fictional Discourse. In: Searle, J. R.: *Expression and Meaning*. Cambridge: Cambridge University Press.

CURRIE, G. (1990): *The Nature of Fiction*. Cambridge: Cambridge University Press.

MARTÍNEZ-BONATI, F. (1981): *Fictive Discourse and the Structures of Literature. A Phenomenological Approach*. Ithaca (N.Y.): Cornell University Press.

OHMANN, R. (1971): Speech Acts and Definition of Literature. *Philosophy and Rhetoric* IV, No. 1, 1-19.

SEARLE, J. R. (1997): *The Construction of Social Reality*. New York: Free Press.

SMITH, B. H. (1978): *On the Margins of Discourse: The Relation of Literature to Language*. Chicago: Chicago University Press.

WALTON, K. (1990): *Mimesis as Make-Believe*. Cambridge, (Mass.): Harvard University Press.

Fictional and Factual Autobiography from the Perspective of Speech Act Theory

Jan Tlustý

Masaryk University, Brno

Abstract: The study considers the pertinence of Searle's speech act theory to literary studies, more specifically to the definition of fictional and factual autobiography. Searle's conception of fictional discourse as a pretended assertion influenced Gérard Genette in his definition of the necessary condition of factual autobiography, which he sees in the identity of the author with the narrator, meaning that in autobiography, the author bears full responsibility for what he or she asserts. According to Genette, if the narrator and the author are not identical, the autobiography is fictional. These theoretical arguments are then applied to the interpretation of Bohumil Hrabal's trilogy *In-House Weddings*, *Vita Nuova* and *Gaps* and the fictional autobiography *Boyhood*, *Youth* and *Summertime* by John Maxwell Coetzee.

Keywords: fictional autobiography, speech act theory, John Searle, John Maxwell Coetzee, Bohumil Hrabal.

Speech act theory is a great source of inspiration for literary studies. Probably the most significant influence can be detected in the field of theory of fiction: even before literary theorists such as Lubomír Doležel, Thomas Pavel and Ruth Ronen came up with their ideas on fiction, literary scholars were intrigued by John Searle's paper *The Logical Status of Fictional Discourse*.[1] This paper stimulated a discussion about whether fiction can be defined semantically or whether it is better to approach it from the perspective of pragmatics and observe what kind of game the author of fiction plays with the reader, that is to ask what the author

---

[1] Searle's study "The Logical Status of Fictional Discourse" was first published in a journal in 1975; later it was included in the book *Expression and Meaning* (see Searle 1979, 58-75).

"does" by writing fiction, what speech acts he performs and how these speech acts are understood by the reader. I will not go into detail of this large discussion which is still very much alive. This study only focuses on one question in the theory of fiction – on the relation between factual and fictional autobiography. I will attempt to show what kinds of stimuli precisely Searle's speech act theory brings to literary studies with respect to this problem and at the same time I will connect this issue to an analysis of specific literary works: I am going to discuss the reminiscent narrative in Bohumil Hrabal's loose trilogy *In-House Weddings* (Svatby v domě), *Vita Nuova* and *Gaps* (Proluky), and the fictional autobiography *Boyhood*, *Youth* and *Summertime* by the Nobel Prize in literature winner John Maxwell Coetzee.

The relation between factual and fictional autobiography was one of the topics dealt with by the French literary theorist Gérard Genette, especially in his book *Fiction and Diction* (1991). To a great degree, Genette's conception draws on Searle's speech act theory, or, more specifically, on the notion of fictional discourse as a pretended assertion.[2] Genette thinks along the following lines: in the case of factual autobiography, the author is identical with the narrator and at the same time with the main character (A = N = C; see Genette 1993, 70, 72-73). This entails that the author bears full responsibility for what he or she asserts: his or her narration hence complies with all the constitutive rules that condition successful realization of the speech act of assertion, which means that the narration must be true, verifiable and sincere.[3] As readers, we understand such an utterance as an authentic utterance about its author.

But what happens if the author is narrating about his or her own life, yet chooses not to be identical with the narrator?[4] This situation

---

[2] Genette takes up Searle's conception of fiction not only in the general chapter on fiction "Arts of Fiction" (Genette 1993, 30-53), but also in the context of his ideas on fictional and factual narrative in the chapter "Fictional Narrative, Factual Narrative" (see Genette 1993, 54-84).

[3] Searle speaks about the essential rule, the preparatory rules and the sincerity rule (cf. Searle 1979, 62).

[4] It is important to remember that Genette does not view the identity of the author with the narrator as identity in terms of name, but he defines it (again, drawing on Searle) as the "author's serious adherence to a narrative whose veracity he assumes" (Genette 1993, 75).

arises when the autobiography is written in the third person. According to Genette, this kind of narrative is necessarily fictional, which is true even if the depicted events accord with the events in the actual world. The author cannot guarantee the veracity of what is said in the text; the whole of the "responsibility" rests on the shoulders of the narrator. I find Genette's use of Searle's theory tremendously inspiring, as it helps us understand experiments taking place in the area of literature and autobiography. It also suggests certain interpretive strategies on the part of the reader. What follows is an attempt to examine this theoretical framework in the context of literature. First I will discuss the reminiscent narratives by Bohumil Hrabal.

Late in his life, in the first half of the 1980s, Hrabal wrote a loose trilogy of recollections: *In-House Weddings* (Svatby v domě), *Vita Nuova* and *Gaps* (Proluky). Hrabal recounts his life in the 1950s and 1960s; he recalls his life with his wife Eliška Plevová (nicknamed Pipsi), his friends, the philosopher Egon Bondy and the visual artist Vladimír Boudník, and the beginnings of his writing. Yet Hrabal's narrative differs from traditional reminiscent narratives: the narrator is not Hrabal, but his wife Pipsi. All the narrated events are thus seen from the perspective of Hrabal's wife. If one applies the speech act theory to this narrative situation, it can be argued that Hrabal pretends to be a fictional character called Pipsi, and as Pipsi he makes an assertion, hence Hrabal imitates the illocutionary act of assertion. As Pipsi's utterance is a figment of Hrabal's imagination, it cannot be subject to rules that usually pertain to assertions. It is thus pointless to ask whether the events really occurred as they are depicted, since the author cannot comply with the constitutive rules of assertion or guarantee the accuracy of the events.

Importantly, however, this does not mean that we cannot consider the potential correspondence between fictional and real events. Genette himself, drawing on Searle and the American literary theorist Barbara Herrnstein Smith, points out that the term fictionality refers to the act of telling itself and not to its content (cf. Genette 1993, 71). When speaking about Bohumil Hrabal, one should remember that his experiment in autobiography tops off his lifelong project of art: to make life the subject of imaginative narration. From the very beginning, Hrabal found inspiration in the reality he experienced or observed, and he also used this reality to create his short stories, novellas and novels. Hrabal certainly does not narrate in order to relate the events of his life to us,

but to shape these events into an art form which will have an aesthetic effect on the reader. In the case of the fictional, reminiscent trilogy, the use of the narrative perspective of Pipsi enables Hrabal to step back and look at himself from a distance, be self-ironic, comment on his own writing and faults, and see himself through the eyes of others. I am going to quote at least a short excerpt:

> Usually my husband met me at the tram stop after my shift at the Hotel Paris Sometimes though he wasn't there which of course meant he was at home in bed because he'd drunk too much and was ashamed of himself That was my husband's thing post-binge shame and crawling into bed so I couldn't see how wobbly he was […]. (Hrabal 2010, 39)

By choosing Pipsi as the narrator of the monologue, Hrabal at the same time offers the fictional character an opportunity to talk about her difficult life and to contrast it with the lives of other people: as a child of German parents, Pipsi was persecuted after World War II and her parents had to leave Czechoslovakia; Pipsi was left behind on her own, marked by the stigma of collective guilt. Through Hrabal's narrative, we look into her mind, discover her worries and fears, and we also follow her transformation after she meets her future husband. It is not necessary to go into detail: I just want to stress that Hrabal offers the reader an opportunity to contribute to the meaning of the novel: for example, in the second book *Vita Nuova* (see the excerpt above), the author does not use any punctuation or paragraphs at all, which encourages the reader to insert some pauses that will lend a certain degree of clarity and hierarchy to the otherwise homogenous text.

Hrabal's reminiscent trilogy can thus be seen as a special case of a fictional autobiographical narrative: the author, even though he narrates about his life, does not bear any responsibility for the veracity of the depicted events. This situation results from Hrabal's unusual narrative strategy that consists in letting his wife Pipsi narrate about his life. In this way, Hrabal made his life into a novel; by refusing to bear the responsibility for the depicted events, he made it possible for the text to have an aesthetic effect on the reader. The special quality of Hrabal's text can be better elucidated by speech act theory, in particular by the conclusions drawn from Searle's theory by Genette (let me mention Genette's idea again: If the author is identical with the narrator, the narrative is factual; if the author is not identical with the narrator, the

narrative is fictional; in this case the narrative is fictional even if the narrated events and facts accord with the events and facts of the actual world).[5] The difference between factuality and fictionality therefore fully depends on whether the author can be held responsible for the represented events, as is the case with autobiography, or whether the assertion is pretended, as it occurs in fiction.

I will now move on to the second example, which is the loose reminiscent trilogy by John Maxwell Coetzee *Boyhood*, *Youth* and *Summertime*. In 1997, J. M. Coetzee published a book of recollections called *Boyhood*, in which he describes the period of his childhood spent in South Africa; five years later he published *Youth* in which he gives an account of his stay in Britain at the beginning of the 1960s. Coetzee, too, tells the story of his life in an innovative way: he does not recount his experiences from the perspective of an "I" and thus there is no "first-person-narrative situation"; he looks at his life from the perspective of a "he." The books do not contain classical reminiscent narration, but loosely arranged scenes which are narrated in the present tense and in which the narrator mediates an insight into the mind of the character called John Coetzee.

Some reviewers have described *Boyhood* as a memoir while mentioning Coetzee's experimenting with the manner of narration. The question is: is this an autobiography or a fictional narrative, hence a novel? Determining the genre of a book is not a purely descriptive endeavour; it influences the reception of the text in a fundamental way: are we to read *Youth* and *Boyhood* as an account of the life of John Maxwell Coetzee, or as a fictional account whose telos is an aesthetic effect? I think that this complex question can be tackled with the help of Genette's ideas about the relation between factual and fictional autobiography: if we were to read Coetzee's reminiscent books as autobiography, the sentences would be subject to rules that apply to ordinary assertions (and the author would be held responsible for these assertions). But (and Coetzee himself is aware of this): talking about the past is tricky, as we are often self-deceived and remember things differently than they actually occurred; we have a natural tendency to forget or alter our past (such as to idealize the past or only recall events that were traumatic). What is then the relation between what really happened

---

[5]   Furthermore, according to Genette, the relation between the author and the narrator is a pragmatic relation (Genette 1993, 78).

and what stays in memory? Can we have genuine access to our own past? Coetzee has dealt with such questions on the theoretical level and he must have been aware of them when writing *Youth* and *Boyhood* as well.[6] In both books he decided to make a similar move like Bohumil Hrabal: he waived his responsibility for the narrated facts and gave the floor to the narrator who is not identical with the author. In accordance with Genette's approach, Coetzee's text can be regarded as fictional autobiography.

However, Coetzee surprised his readers again in 2007 when he published *Summertime*, the last book in the trilogy. While in the first two books the reader could still hesitate regarding the relation of fiction and autobiography, in the third one the author sends a clear message about the fictitiousness of the narrative: one of the characters is a biographer Vincent who is writing a monograph about the deceased writer John Maxwell Coetzee. For the purposes of his monograph, Vincent goes to see various people who played an important part in Coetzee's life: their narration is a part of the text and thus the reader can piece together a colourful picture of Coetzee's life. What Coetzee seems to be saying in this novel is that grasping one's life, as well as the life of others, is complicated, and fraught with peril, misunderstanding and misinterpretation. Is the story of our life the story that *we* tell about ourselves, or is it the story that *others* tell about us? I think that the activation of these questions – this fictional research into autobiographical narration – constitutes an important part of the aesthetic effect of Coetzee's fictional memoir. The example of his own life serves as an illustration of Coetzee's inquiry into what is human memory, what dangers there are in writing an (auto)biography and whether it is at all possible to get to know oneself or another person.

Not only can the fact that Coetzee's first two books are fictional, that is pretended autobiography, inferred from Genette's theory, it is also confirmed by Coetzee himself when he uses the same subtitle "Scenes from Provincial Life" in *Summertime* as in the previous two books. From the point of view of reception, this is an unusual situation because the novel *Summertime* can retrospectively influence the reading of the two previous books, as it sends an unambiguous message that *Boyhood* and

_____

6   Coetzee has discussed the issue of autobiography in his inauguration lecture at the University of Cape Town (see Coetzee 1984, 1-6) and in the chapter "Autobiography and Confession" in his *Doubling the Point* (see Coetzee 1992, 241-293).

Youth are fictional. As if to corroborate this argument, all three books appeared in one volume called *Scenes from Provincial Life* in 2011, which accentuates the fictionalization of the autobiographical narrative even further.

In this paper, I have attempted to demonstrate that speech act theory can be methodologically relevant and inspiring to literary studies, in particular to analyses of the relations between factual and fictional autobiography. Conclusions drawn from Searle's theory by Gérard Genette illuminate the ontological status of both types of narratives, but they also enable us to describe the readerly competencies that these texts expect their readers to possess.

Department of Czech Literature and Library Studies
Faculty of Arts
Masaryk University
Arna Nováka 1
602 00 Brno
Czech Republic
tlustyj@email.cz

References

Coetzee, J. M. (1984): *Truth in Autobiography*. Cape Town: University of Cape Town, Inaugural lecture, new series, no. 94.

Coetzee, J. M. (1992): *Doubling the Point: Essays and Interviews*. Cambridge (Mass.) – London: Harvard University Press.

Coetzee, J. M. (1997): *Boyhood: Scenes from Provincial Life*. London: Secker & Warburg.

Coetzee, J. M. (2002): *Youth: Scenes from Provincial Life II*. London: Secker & Warburg.

Coetzee, J. M. (2009): *Summertime: Scenes from Provincial Life*. London: Harvill Secker.

Genette, G. (1993): *Fiction and Diction*. Ithaca and London: Cornell University Press.

Hrabal, B. (2007): *In-House Weddings*. Northestern: Northwestern University Press.

Hrabal, B. (2010): *Vita Nuova*. Northestern: Northwestern University Press.

Hrabal, B. (2011): *Gaps*. Northestern: Northwestern University Press.

Searle, J. (1969): *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

Searle, J. (1979): *Expression and Meaning.* Cambridge: Cambridge University Press.

Searle on Metaphor

Jakub Mácha

Masaryk University, Brno

Abstract: The main aim of this paper is to survey and evaluate Searle's account of metaphor (1979) in the light of Davidson's arguments against the idea of metaphorical meaning, which appeared at roughly the same time. Since this paper is intended for a festschrift celebrating Searle's respectable anniversary, I will mostly refrain from critical remarks and rather focus on the positive aspects of his account. I am going to show that Searle's theory of metaphor is for the most part immune to Davidson's arguments.

Keywords: metaphor; Searle; Davidson; literal meaning; metaphorical meaning.

Let me introduce the problem in question with an example of a rich metaphor by Wallace Stevens. He wrote the following verse-lines in his poem (2006, 60) "Sunday Morning":

Death is the mother of beauty; hence from her,
Alone, shall come fulfillment to our dreams
And our desires.

To say that death is the mother of beauty is literally false. One may ask, then, what this metaphor means or in what sense it could be true. One may ask whether the poet intended to communicate some definite insight. One can fend off these questions by pointing out that such questions would deprive us of all poetic effects. The poet himself, however, infers other nontrivial insights about our dreams and desires from the metaphor. Hence, he might have meant something definite by the metaphor. These questions and considerations express intuitions behind the theories of metaphor that I am going to focus on.

1   Searle's account of metaphor

Searle builds his account of metaphor on his speech act theory based on a general Gricean framework. The main question he aims to investigate is "How do metaphorical utterances work, that is, how is it possible for speakers to communicate to hearers when speaking metaphorically inasmuch as they do not say what they mean?" (Searle 1979, 92). Since speakers mean and try to communicate something other than they say, metaphorical utterances are, thus, indirect speech acts. To be more formal, the speaker says that S is P and means metaphorically that S is R. Searle calls what the speaker says (that is "S is P") *a sentence meaning* and what she means (that is "S is R") *the speaker's utterance meaning*. The question is, thus, how it is possible to say "S is P" and both mean and communicate "S is R" where P does not literally mean R. Searle put this argument this way: If one can communicate "S is R" using "S is P" then the relation between the sentence meaning and the utterance meaning must be systematic. The next task is now to find shared principles or strategies of how to arrive from the sentence meaning to the utterance meaning.

First, there must be a principle that allows that the speaker's utterance will be taken metaphorically and that the hearer will recognize that the utterance is not meant literally, but metaphorically. The most common strategy is here to check out whether the utterance is obviously defective if taken literally, i.e. whether it is patently false or true.

Second, there must be principles generating all the possible values of the R term from the P term. Searle admits that there is no single principle that is distinctive about metaphorical utterances. He lists eight principles with the suspicion that there might be even more. Let me quote these principles and their examples from a compendious survey in Camp (2003, Ch. 1.2):

1. Things which are P are by definition R; usually R will be one of the salient defining characteristics of S. Example: "Sam is giant" means$_{met}$ "Sam is big".
2. Things which are P are contingently R; again, R will usually be a salient or well-known property of P things. Example: "Sam is a pig" means$_{met}$ "Sam is filthy, gluttonous, and sloppy, etc."
3. Things which are P are often said or believed to be R, even though both speaker and hearer may know that R does not in fact apply to

P things. Example: "Richard is a gorilla" means_{met} "Richard is mean, nasty, prone to violence, and so on."

4. It is a fact about our sensibility, whether culturally or naturally determined, that we just do perceive a connection, so that P is associated in our minds with R. Example: "Sally is a block of ice" means_{met} "Sally is unemotional."

5. The condition of being P is like the condition of being R. Example: "You have become an aristocrat" means_{met} "Your new status is like that of being an aristocrat".

6. P and R are the same or similar in meaning, but one, usually P, is restricted in its application, and does not literally apply to S. Example: "His brain is addled" (no interpretation provided).

7. A principle extending the simple 'S is P' form to other syntactical forms, basically by applying 1-6 at a higher order. Example: "The ship ploughs the sea" means_{met} "The ship moves the sea to the side of the prow as it moves forward".

8. P and R may be related as part-whole or container-contained, so that metonymy and synecdoche also count as metaphors.[1]

Notice that these principles are not context-dependent nor are their input data taken from the context. The only context-dependence here is based on the fact that the meanings of P and R as such are context-dependent, as will be explained later. The principles simply relate the predicate terms P and R as though they stood in isolation and thereby generate all possible values of the R term.

Third, the range of the possible values of R has to be restricted to the possible properties of the subject term S. Here the context enters again (the first step is context-dependent too) as it is a matter of the context that the predicate P, which is the basis for generating the possible values of R, stands in a predicative sentence together with the subject S.

Finally, I want to mention a feature that is distinctive of Searle's account and that has gone unnoticed even by authors sympathetic to him. Before introducing the principles presented above, Searle aims at a characterization of literal utterances. This is an extremely difficult task, for it amounts to characterizing predication in general. However, if we did not have an account of literal utterances and of literal meaning (at

---

[1] The notation 'means_{met}' abbreviates 'can be uttered metaphorically to mean that'.

least in a rough outline), then claims about metaphorical meaning (e.g. that it is secondary to the literal meaning or that there is no metaphorical meaning) would not say much. So, in a literal utterance, as already mentioned, the sentence meaning and the utterance meaning coincide. And, more important, "the literal meaning of a sentence only determines a set of truth conditions relative to a set of background assumptions which are not part of the semantic content of the sentence"(Searle 1979, 96). It is also futile to say that metaphorical meaning is context-dependent or open-ended, because literal meaning could have these features as well.

2 Davidson's arguments against the of idea of metaphorical meaning

It is probably only a coincidence that Davidson's paper "What Metaphors Mean" appeared in the same year as Searle's paper "Metaphor"; so, I presume that there was no mutual influence. Davidson's paper is mostly critical in focus, attacking the semantic theories of metaphor, especially the one given by Max Black (1955). Davidson's main claim is that "metaphors mean what the words, in their most literal interpretation, mean, and nothing more" (Davidson 2001, 245). Hence, Davidson denies that metaphors have any figurative, second or metaphorical meaning. "The central error about metaphor is most easily attacked when it takes the form of a theory of metaphorical meaning, but behind that theory, and statable independently, is the thesis that associated with a metaphor is a definite cognitive content that its author wishes to convey and that the interpreter must grasp if he is to get the message" (Davidson 2001, 262). We can read off two claims from this quotation. First, metaphors have no metaphorical meaning, and second, metaphors do not serve as means of communication. Both claims are in an apparent contradiction to Searle's view, so we have to look carefully at Davidson's intuitions and arguments for the support of his ideas.[2]

One thing has to be pointed out at the outset. Davidson uses the expression "meaning" without any qualification. One could wonder whether his arguments are valid for any conception of (literal) meaning

---

[2]   I tried to give an exhaustive list and critical discussion of Davidson's arguments in my book Mácha (2010). Other philosophers, e.g. Reimer (2001) or Lycan (forthcoming), offer other expositions of some of these arguments.

or whether they are restricted to his own minimal account of meaning and interpretation. I will first suppose that the former option is more plausible, since otherwise any reference of other philosophers' views would be impossible. Davidson's main thesis is, thus, that there is no metaphorical meaning beyond the literal one *in any conception of literal meaning*.

Intuition 1: There are no rules for the construction of metaphorical meaning or metaphorical content. "There are no instructions for devising metaphors; there is no manual for determining what a metaphor 'means' or 'says'; there is no test for metaphor that does not call for taste" (Davidson 2001, 245). Davidson offers no justification of this claim. It is rather an intuition of his, stated at the outset of his paper. Producing and understanding metaphors is a creative endeavor. If it were bound by rigorous rules, then the construction of metaphorical meaning would be a mechanical process and all creativity would be lost. I think something like this lies behind this intuition. Taken the other way around, if metaphor counted as a semantic or pragmatic phenomenon, there would have to be such rules. Davidson's intuition is, hence, that metaphor is neither a semantic, nor a pragmatic phenomenon.

Reply: Searle does not share this intuition with Davidson. His intuition or presupposition is that metaphors could be used as means of communication. It follows that there must be rules for construction of metaphorical meaning. Metaphor is hereby located among pragmatic phenomena and could be explained within the framework of Searle's speech act theory. The rules Searle gives are a plausible theoretical reconstruction of our understanding of metaphors. But the issue of creativity remains. Do these rules pose a problem for the claim that producing and understanding metaphors is creative? I think they do not. These rules are not formulated strictly. They allow a sort of creative freedom. Rule 4, for example, evokes a culturally or naturally determined association between literal and metaphorical meaning. To find out, however, which association exactly is the case, requires a lot of creativity. Similarly, rule 5 is based on the conditions of being P and R. But which condition exactly is the case is left to the speaker's and addressee's cooperative effort.

Argument 2: The idea of metaphorical meaning does not explain how metaphors work. "These ideas don't explain metaphor, metaphor explains them. Once we understand a metaphor we can call what we grasp the 'metaphorical truth' and (up to a point) say what the 'meta-

phorical meaning' is. But simply to lodge this meaning in the metaphor is like explaining why a pill puts you to sleep by saying it has a dormative power" (Davidson 2001, 247). The idea of literal meaning has explanatory power, because literal meaning can be assigned to words (or linguistic items) apart from particular contexts of use. We cannot do the same with metaphorical meanings. Davidson's argument is, hence, that since we cannot assign metaphorical meanings apart from the contexts of use, the idea of metaphorical meaning has no explanatory power, and it is also pointless to postulate such a superfluous thing. If we could assign a metaphorical meaning to a metaphor regardless of the context of use, then the metaphor would become a dead one. The metaphorical meaning could then be a second meaning and could be enlisted in a lexicon.

Reply: This argument is directed against the semantic accounts of metaphorical meaning. Searle's metaphorical meaning is not a sentence meaning, but an utterance meaning which is assigned only in the context of use. Since there are principles stating how to generate a metaphorical meaning out of a (literal) sentence meaning and of shared background assumptions, metaphorical meaning is endowed with genuine explanatory power.

Argument 3: Dead metaphors involve literal meanings, but these are not fossilized metaphorical meanings. The first premise of this argument is that "If metaphor involved a second meaning, as ambiguity does, we might expect to be able to specify the special meaning of a word in a metaphorical setting by waiting until the metaphor dies" (Davidson 2001, 254). Davidson argues then that literal meanings are usually poor, simple or narrow compared with the way metaphors work. Thus, literal meanings of dead metaphors cannot be based on metaphorical meanings of living metaphors. Hence, there are no metaphorical meanings.

Reply: On Searle's account, literal meaning is not as narrow as Davidson's argument requires. The dying of a living metaphor may be explained as the settling down of one of its utterance meanings that happens to be so common that it could be assigned independently of any context of use and, thus, becomes a second sentence meaning. This argument, in fact, restricts the validity of Davidson's critical remarks on conceptions of meaning that are minimalistic like his own theory of meaning and interpretation or the semantic minimalism of his followers (see Lepore – Stone forthcoming).

Argument 4: If similes do not have a second meaning, then neither do metaphors. I do not want to question the premise. This argument presupposes, however, that metaphors and similes share the same logical form. Certainly, both figures involve similarities between their subjects. But why should they share the same logical form? It is not *prima facie* clear what the logical form of simile *A is like B* is. Is it an existential statement that there is a similarity between A and B or is it an assertion of a contextually salient similarity between the terms? In the former case, similes are trivial and there is no problem that the speaker literally means what she says by the simile. In the latter case, similes are figurative in the same way as metaphors are, and one has to provide an account of how to determine the similarity in question. In this case, similes and metaphors share the same logical form except that most similes are true and most metaphors are false. This could be the reason for postulating a second meaning to metaphors as opposed to similes. Following Lycan (forthcoming) I find this argument entirely unconvincing.

Reply: Searle (1979, 103) maintains that similes could be figurative and so they "need not necessarily commit the speaker to a literal state-ment of similarity". The speaker of a simile could mean something dif-ferent from what he says just like in a metaphor.

Argument 5: If metaphors had second meanings, these would be pos-sible to express in literal paraphrases. But metaphors are, in general, not amenable to literal paraphrases. Hence, metaphors do not involve second meanings. Davidson offers a reason why this is so: "If what the metaphor makes us notice were finite in scope and propositional in na-ture, this would not in itself make trouble; we would simply project the content the metaphor brought to mind on to the metaphor. But in fact there is no limit to what a metaphor calls to our attention, and much of what we are caused to notice is not propositional in character. When we try to say what a metaphor 'means', we soon realize there is no end to what we want to mention" (Davidson 2001, 262). This is to understand that some metaphors are not amenable to literal paraphrase, because their content, i.e. what they bring to mind, cannot be delimited. Some metaphors are open-ended. Again, there are several unarticulated as-sumptions: First, literal meanings and metaphorical meanings are of the same kind, and second, literal language is not open-ended in the same way as metaphors are.

Reply: The open-endedness of metaphorical meaning is guaranteed by the fact that it is derived from literal meaning which is, on Searle's account, open-ended as well. That all metaphors are amenable to literal paraphrases follows from his Principle of Expressibility which can be summarized as "whatever can be meant can be said" (cf. Searle 1969, 17). If the speaker means and wishes to communicate something by a metaphor, then it can be expressed in a literal paraphrase. Davidson's counterargument may be that the speaker might intend to do something other, namely to induce an indefinite (perlocutionary) effect. If so, then the metaphor would not be counted as a pragmatic phenomenon. This is a serious problem for Searle's account, which concerns, however, his basic assumptions rather than his arguments. In the same vein, one could side with Searle and dismiss Davidson's intuition that metaphor belongs to the perlocutionary realm.[3]

Argument/intuition 6: This argument is based on the intuition that genuine metaphors can be appreciated repeatedly without losing their metaphorical nature. "Novelty is not the issue. In its context a word once taken for a metaphor remains a metaphor on the hundredth hearing, while a word may easily be appreciated in a new literal role on a first encounter. What we call the element of novelty or surprise in a metaphor is a built-in aesthetic feature we can experience again and again, like the surprise in Haydn's Symphony No. 94, or a familiar deceptive cadence" (Davidson 2001, 252). Davidson's intuition is that we could read or hear a metaphor *in the same context of use* repeatedly and its effect might be different on each occasion. If this metaphor had a second meaning (although derived from this context), the effect of this meaning would be always the same. Hence, metaphors have no second meanings.

Reply: If we accepted this intuition, we could ask whether a sentence has the same utterance meaning if it is read or heard repeatedly in the same context. To derive an utterance meaning amounts to identifying the "possible speaker's intentions" (Searle 1979, 93). Utterance meaning

---

[3] Lycan (forthcoming) offers a sort of rapprochement of Searle's and Davidson's accounts. He argues that there is a continuum between metaphors that could be explained pragmatically and metaphors whose point is a perlocutionary effect. It has to be mentioned, however, that although in general sympathetic to Searle's account, he nevertheless thinks that the open-endedness poses a problem here.

is also derived from what the speaker might mean by her utterance, not necessarily from that what she actually means. This might be different in different occasions due to the fact that the shared background assumptions may be different.

To have a complete survey of the controversy in question, let me sketch Davidson's positive account of metaphor, since he left nothing more than a sketch. As already indicated, Davidson sees metaphor as a perlocutionary effect that cannot be explained within semantics or pragmatics. The point of a metaphor exceeds any regular (i.e. rule governed) comprehension. What can be made out of this conviction? Rorty (1987) developed Davidson's views in the way that an explanation of our comprehension of metaphors can be given only in terms of causal connections and psychological associations. But Davidson, following Wittgenstein, also claims that a metaphor lets us *see* one thing *as* another thing. Then what metaphors let us notice is not propositional in focus but nevertheless can be explained as a formal relation between two concepts. And that is much more than a causal effect. Making use of Wittgenstein's analysis of the phenomenon of seeing-as, I have argued in Mácha (2010, 136-142) that in a metaphor "A is B" one can perceive and think of an internal relation between the concepts A and B.

3 An assessment and possibly an improvement of Searle's account

The clash exposed in the previous section can be seen as a clash of intuitions rather than arguments. But the authors that try to develop Searle's account of metaphor further have also raised several objections. The most important one which I am going to address here is that the principles of generating a metaphorical meaning are too vague and not distinctive of metaphor.[4] The following quotation from Camp (2003, Ch. 1.2) is characteristic: "each of the principles adduced is itself so broad, and the list as a whole comprises so many different ways in which *P* and *R* might be related, that in the end they amount to not much more than the requirement that *P* and *R* must be similar (or just

---

[4]   Davidson (1979, 262) makes this point in general: "It should make us suspect the theory [of metaphorical meaning] that it is so hard to decide, even in the case of the simplest metaphors, *exactly* what the content is supposed to be" (my emphasis).

related) in some respect or other." These principles cannot be, on the other hand, too rigid, as argued in the reply to Davidson's intuition 1 above. This creates a tension in Searle's account. Camp further argues that the eight principles generate an indefinitely long list of features. These features have to be restricted in the third step to those that are possible features of the metaphorical subject S.

My worry is, then, that due to the vagueness of these principles, all possible features of S will be generated. In each case, these principles could generate a *possible* feature of S that S *actually* has. The upshot of this argument would be that all metaphors are (analytically) true and hence not capable to communicate anything.

If a metaphor should carry some distinctive message about its subject S, this content has to be delimited. It has to be set out what the content is and *what it is not*. The third step of Searle's construction of the metaphorical meaning gives, however, no clues in this respect. Let us take the metaphor "Sam is a pig". Among the features of R generated in the second step might be "greedy" and "slovenly". They are both among the possible features of a person; hence, they will pass through the third step. Then we are left uncertain of the metaphorical meaning. Is it that Sam is a greedy person or that Sam is a slovenly person?

What we also need is a principle that would restrict the features generated in the second step even more. This principle has to delimit which possible features of the metaphorical subject S are parts of the metaphorical meaning and which are not. Only then would the metaphor be capable of being true or false and hence capable of communicating a cognitive content. I have no such principle at hand. However, my suspicion is that it has to take into account more information from the context of use and from the shared background assumptions. If so, then processes of pragmatic inference would contribute to the truth-conditional content of a metaphor. Then we would reveal something like *Grice's Circle* (see Levinson 2000, 186) in the case of metaphors.[5]

There is another objection to Searle's account of metaphor that I would like to address here. It is that his principles are not distinctive of metaphor. They may apply to other figures or indirect speech acts and implicatures. If we leave aside the eager effort to find the essence of

---

[5]   Consider the metaphor "Death is the mother of beauty" again. Its utterance meaning depends on the implicature "from her, alone, shall come fulfill-ment to our dreams and our desires."

a phenomenon, the non-distinctiveness of the principles could be seen rather as an advantage. Searle gives a single universal mechanism that can be applied when interpreting other figures and linguistic phenomena in general.[6]

It is a matter of philosophical taste whether one wants to have a robust theory of meaning capable of explaining various non-standard phenomena or whether one can strive for a minimalistic theory of meaning that leaves all anomalousness outside.[7] An in depth discussion of this topic, however, exceeds the scope of the present paper. Davidson's arguments are valid only for the minimalistic conception of (literal) meaning. If one takes a richer account of meaning, as in Searle or in the contemporary contextualism (see, e.g., Bezuidenhout 2001; or Recanati 2001), metaphor can be interpreted as a pragmatic or even a semantic phenomenon.

Departmenf of Philosophy
Faculty of Arts
Masaryk University
Arna Nováka 1
602 00 Brno
Czech Republic
macha@phil.muni.cz

References

Bezuidenhout, A. (2001): Metaphor and What is Said: A Defense of a Direct Expression View of Metaphor. *Midwest Studies in Philosophy* 25, 156-186.
Black, M. (1955): Metaphor. *Proceedings of the Aristotelian Society* 55, 273-294.
Camp, E. (2003): *Saying and Seeing-as: The Linguistic Uses and Cognitive Effects of Metaphor*. PhD. Dissertation, UC Berkeley.

_____

[6] Camp (2003, Ch. 1.2) reports a personal communication with Searle concerning this objection: "The fact that the principles he [Searle] adduces apply not just to metaphor, but to indirect speech and implicature more generally, should not be seen an objection to his account but as a further advantage of it, he says."

[7] An effort of locating metaphor outside a theory of meaning is expressed in Lepore – Stone (forthcoming): "Progress, we have argued […], depends on developing and applying clear standards that demarcate semantic phenomena narrowly. If we can locate metaphor elsewhere, it is good news for meaning."

DAVIDSON, D. (1979): What Metaphors Mean. In: Sacks, S. (ed.): *On Metaphor*. Chicago: University of Chicago Press. Reprinted in Davidson, D. (2001): *Inquiries into Truth and Interpretation*. Second Ed., Oxford: Clarendon Press, 245-264. (All citations in the text are from the reprint.)

LEPORE, E. – STONE, M. (forthcoming): Against Metaphorical meaning. *Topoi*.

LEVINSON, S. (2000): *Presumptive Meanings: The Theory of Generalized Conversational Implicature Language, Speech, and Communication*. Cambridge, (Mass.): The MIT Press.

LYCAN, W. (forthcoming): An Irenic Idea about Metaphor. *Philosophy*.

MÁCHA, J. (2010): *Analytische Theorien der Metapher*. Berlin: LIT Verlag.

RECANATI, F. (2001): Literal/Nonliteral. *Midwest Studies in Philosophy* 25, 264-274.

REIMER, M. (2001): Davidson on Metaphor. *Midwest Studies in Philosophy* 25, 142-155.

RORTY, R. (1987): Unfamiliar Noises I: Hesse and Davidson on Metaphor. *Proceedings of the Aristotelian Society*. S*upplement volume* 61, 283-296.

SEARLE, J. (1969): *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

SEARLE, J. (1979): Metaphor. In: Ortony, A. (ed.): *Metaphor and Thought*. First Ed., Cambridge: Cambridge University Press.

STEVENS, W. (2006): *Collected Poems*. London: Faber and Faber.

Reply to Commentators

John Searle

I am immensely grateful that so many people have commented on my work, and I will attempt to make at least a brief reply to each commentator. I want to use this preface to the replies by apologizing to everyone that my answers are really inadequate, in large part because I do not have the time to go into detail for fourteen different sets of comments, some of which were quite extended and complex. There is no common theme that pervades all of the articles, so I will simply deal with each one as it was presented to me. Philosophy is by its nature an argumentative discipline, and I follow the tradition in emphasizing points of disagreement more than agreement, but this does not diminish my gratitude for the thought and effort that went into creating the works I disagree with.

I.

Lukáš Zámečník: External Realism as Non-Epistemic Thesis

Zámečník is right in understanding my conception of realism as not a hypothesis along with others, but rather as a condition of the intelligibility of a large class of representations such as statements and beliefs. He refers to a lot of authors that I have not read and refers to something persistently as the "analytical tradition". I doubt if there is anything like a unified "analytical tradition" on this issue.

I am very unsympathetic with Quine's conception of naturalized epistemology, at least as far as I understand it. I think Zámečník regards his thought experiment with Luke and Saman as important, but I really did not understand it; so I am not quite sure what significance it is. He is right to think that, polemically, my arguments about external realism were aimed at certain fashions. They were aimed at more or less incoherent views, such as post-structuralism and post-modernism.

However, that is not my only motivation. I think it is an interesting philosophical question 'What exactly is the status of propositions such as that there exists a reality independent of representations of it?' He is right that I think of ER not as theory among others, but rather as a condition of a certain kind of intelligibility.

Tomáš Marvan: Searle on External Realism and "Privileged Conceptual Scheme"

I am grateful to Tomáš Marvan for raising so many important issues about external realism and conceptual relativism. I think he misunderstands my view in fairly radical ways, and my main objective in this response would be to correct his misunderstandings. He thinks that somehow on my view of the world it is featureless; it just comes as a "featureless, shapeless lump" (p. 35). He says: "Either the world contains dogs, grass and supernovae without humans and their representation, or it doesn't. If it does, we cannot divide it up in any way we please – provided our aim is to describe it correctly"(p. 35).

Of course the world contains dogs, grass, and supernovae without humans, but what it does not contain without humans is these *labels* and *categories*. The labels that enable us to divide the world into dogs, grass, and supernovae are human creations. They are not arbitrary, because we want our categories to match such natural distinctions as spatio-temporal boundaries and causal relations. But my main point is that the world does not come already labeled; we have to invent the verbal and other sorts of categories for describing it.

If you think that, somehow or other, our categories are inevitable; then imagine that we were different sorts of creatures altogether. If you imagine that we typically had the size of galaxies, we would probably not be very interested in giraffes. Giraffes would not cease to exist, but we would not have a category for describing giraffes. Similarly, if we were the size of hydrogen atoms, I doubt very much that we would be interested in planets. The point is that the invention of the categories is up to us, whether or not they apply to reality is up to reality, and we want to get categories that fit our interests in reality. But if we were different sorts of beings with different sorts of interests, we would have different categories.

I can summarize my views in a nutshell by saying that conceptual relativism does not imply metaphysical or ontological relativism. We

adopt the categories that are relative to our interests, but the world does not give a damn about our interests; and whether or not the categories we adopt fit the world or fail to fit the world is up to the world, not up to us. There are a very small set of categories where we create a reality to fit the categories by using expressions that name the categories. These are what I call "institutional facts", and I have attempted to analyze how they are socially constructed in some detail. But it should be obvious that the world as described by physics and chemistry is not in that way socially constructed.

In my conception of realism, it is neutral about how the world actually turns out. It could turn out that we are mistaken in our existing scientific theories, and yet externalism according to me would still be correct. Indeed a necessary presupposition of rationality. It is a mistake to think of realism as a specific view about how the world is. Realism can be right, and yet we can be wrong in every detail of our present views about nature. What is the difference between the view of realism that says, "There is a specific way that the world has to be" and my view that says, "There is just a way that it is and we can be mistaken about specific details"? Why not adopt a more robust version of realism that is committed to specific thesis about how the world is? Why not a version of realism that says it is a part of metaphysical realism that the atomic theory of matter is true? The problem with this is that it cannot account for what is preserved if we discover that atomic theory is wrong. The point is not just a point about the tentativeness of our claims, but rather a point about the nature of representation in general. It is sometimes said that the argument for realism is that scientific theories tend to converge, but the problem with that is that if they did not converge, that fact would have to be discovered as well. Non-convergence is as much an argument for realism as is convergence precisely because a certain Background presupposition is common to both cases. That common Background presupposition is what I call realism.

I think my view of realism is correct and the other is mistaken. On my view, realism is a Background presupposition about the nature of representation in general. It is a much more powerful and important view than just the world happens to be a certain way.

He says there is no way to distinguish my view from constructivism, but that again I believe is mistaken. There is a crucial distinction, because the constructivist denies that there exists a world independently of our representations; whereas I am making a much stronger claim

that the very notion of representations of a certain kind presupposes precisely that the world exists independently of our representations.

Why is it important to make a distinction between a more formal minimalist conception of realism and a more substantive robust concept, and to defend the minimalist rather than the robust? One way to put the answer is this: if somebody wants to defend a robust realism, then he is required to say *exactly* what are the entities, features, etc. that his theory is committed to. What exactly does realism involve? The problem however is that none of the realists he cites, as far as I know, has been willing to do that; and in such a case, as we will see, there is a good reason why they have been unable to do that. And that is, whatever entities they specify as essential, you need to be able to describe a situation in which they might be mistaken about the existence of those entities. Yet something important is preserved in both the cases where, for example, the atomic theory of matter is true and the case where it is not true. That important something is what I call external realism.

Let me conclude this discussion with an example that illustrates the tentativeness of any stage of our scientific knowledge. I have been telling my students literally for decades that the world consists of entities we find it convenient (if not entirely accurate) to call physical particles. But it now turns out that our old friends the "physical particles", molecules, atoms, electrons, etc., are in fact only 4% of the world; the other 96% consist of dark energy and dark matter. What is "dark" about dark matter and dark energy is epistemic darkness. We do not know what is going on. Suppose it turns out that we were wrong in describing our comfortable 4% in the vocabulary of atoms, molecules, and subatomic particles. Suppose the 4% really are special cases of dark matter and dark energy, and there is not anything of the kind that we are familiar with. That seems to me a distinct possibility, and all of this is consistent with external realism as it really matters. The thesis is that there is a way that things are, and our task is to find out how things are. I think he totally misunderstands my view, and he thinks somehow I am committed to what he calls an "amorphous lump".

I can summarize my objections to his paper by saying that I believe he is mistaken in two crucial respects in his conception of my view. I do not think the world is an amorphous lump. It has exactly the same shape it always had. Secondly my view is not a variant of constructivism. It is strongly opposed to constructivism.

I think one of those cases in philosophy, what appears to be dis-agreement is not genuine disagreement. I think in the end we probably agree, and I have not succeeded in making my views sufficiently clear.

Vladimír Havlík: Searle On Emergence

Havlík discusses my conception of emergence. I draw various anal-ogies. For example, I say, roughly speaking, consciousness is to the brain as liquidity is to water. Of course, with all analogies there are limits; and in this case, for example, there are important disanalogies. Liquidity is ontologically reducible to molecular behavior in a way that consciousness is only causally reducible, but not ontologically reduc-ible, to brain processes. All the same, I think the analogy is useful in getting us to see that there is an utterly harmless sense of "higher level" or "emergent" in which consciousness is a higher level or emergent property of the brain. Havlík says that he finds ambiguities in my no-tion and that it is "loaded with a form of mechanicism" (p. 43). I am not sure what the problem is supposed to be.

I opposed the concept I described as emergent #2. What I had in mind were those early theories of consciousness as an emergent prop-erty of the brain which were designed somehow to preserve free will. The brain has consciousness as an emergent property, but once emerged consciousness has powers that cannot be explained by the powers of the brain. This violates the principle of transitivity of causation, and I think there are no convincing examples of emergent #2.

I did not understand his puzzlement about how a system property could fail to be "*the causal consequence of interactions among constitutive entities*" (p. 44). But I think there are obvious cases of this. There are system properties that are not *causal* consequence of the behavior of the elements, for example, weight. The weight of the whole system is arrived at by summation of the weight of the components. But the re-lationship is not causal, it is just additive. For example, if I put a bunch of bricks on one side of a scale, they will force the scale down and thus have a causal power. But the causal power is not a causal consequence of the behavior of the microstructure; it is just a matter of adding the weights of the elements. In the passages he cites I wanted to distinguish properties such as weight and shape – which could be just figured out, so to speak, arithmetically – from those that involve causal interactions

among the components. I would understand his claim better if he gave some examples.

The point I made about deduction is that if you just describe the neurons in terms of their anatomy, you cannot deduce the presence of consciousness. You need some account of the causal relations between them. Given an appropriate account of the causation you could deduce that the system is conscious. So, for example, if there were laws that determined that a system is conscious under certain circumstances, then the description of the circumstances together with the statement of the laws would enable the deduction that the system is conscious.

He thinks there is something mysterious about my use of the expression "additional causal interactions". If you just think of the brain as a collection of neurons, and you do not consider the causal relations among the neurons – for example, neuron firings at synapses – then you would not be able to deduce that the system is conscious. To get to consciousness, you have to know about "additional causal interactions". He calls this a "mysterious incantation". But I think it is pretty clear what it means, and in fact I do try to describe it in detail in various writings. The synapse is identified anatomically. But its role is functional. It is the point at which there is a transmission of a signal from the axon of one neuron to the dendrites of the next neurons in line. Neuron firings at synapses are among the causal relations I am describing.

He leaves out any serious discussion of examples, so let me introduce an example that will clarify things. At one point, Francis Crick proposed that consciousness might be caused by synchronized neuron firings in the range of 40-70 Hz between the thalamus and layers 4 and 6 of the cortex. It turned out that this account was not right, but something like it has to be right; that is to say, there has to be some set of causal relations among neurons that causally accounts for consciousness. To repeat, you cannot deduce consciousness from a description of the anatomy of the neurons, but if you add further accounts of the mechanisms involved and those accounts really do explain consciousness, then you can deduce the presence of consciousness from these explanatory mechanisms. My point was not that consciousness is "non-deducible" just like that; rather it is non-deducible without some additional premises. It is task of neurobiology to provide us with those premises.

One of his most important claims is that my conception of emergence includes a form of mechanicism. I am not sure what mechanicism is. But, in any case, he is certainly right to point out that entities

can alter when they take part in the creation of a system – think of the changes involved in the chemical bond. But I am not sure how this bears on my discussion. He seems to think that, somehow or other, I am committed to the view that entities engaged in causal interactions never alter their character as a result of their causal interactions. That seems to me an absurd view, something I have never defended. But in any case, the question to what extent do neurons alter in the course of their interactions with other neurons is not a philosophical issue. It is a straight forward scientific question to be settled by neurobiological research. The whole point of my discussion was that the anatomy of the neuron is not sufficient to imply consciousness. You have to include a whole series of causal relations among large systems of neurons to be able to account for consciousness.

I am grateful to him for his thoughtful paper. But I think he misunderstands my views in some fairly fundamental ways. Specifically, he misunderstands the claim that consciousness is not deducible from a description of the neurons. The point is: it depends on how much you add to the description. And secondly, he misattributes to me the view that micro entities do not alter during causal relations with other micro entities. This view, we know independently, is false. In any case, it is not a view I have ever defended.

II.

Martin Pokorný: Sentience, Awareness, Consciousness

I share with Pokorný the assumption that consciousness is a biological process like digestion or photosynthesis and as such requires a biological explanation. Such an explanation will cite the causal mechanisms that produce the phenomenon. He seems to think that it is necessary to argue for this point. I take it for granted. Like a lot of authors, he thinks complexity is somehow relevant or essential to consciousness. This may be so, but I have never seen an argument for it. Simple forms of consciousness, for all we know, may be produced by simple mechanisms.

The problem of consciousness in its simplest form is to explain how mechanisms such as the brain that have an objective or third person ontology can cause (produce, give rise to, result in) phenomena that have a subjective or 1st person ontology. This is the problem of explaining how brains cause consciousness. The problem is, How can brain processes described in objective third person terms produce something

that feels like something? For every conscious state, there is something that it feels like to be in that state. The question is, How do such states come out of processes that have no feelings? The account of Martin Pokorný not only fails to solve this problem, he does not even address it or seem aware of its existence. All of the mechanisms that he cites are third person mechanisms, and he does not tell us how they are supposed to produce a first-person ontology. He places heavy reliance on the notion of recognition as defined by Gerald Edelman. However, this is entirely a third person process. Let me emphasize this: there is no psychological reality at all to the notion of recognition so defined. And it is the total mystification to suppose that it could somehow produce subjectivity without any further explication. He introduces two notions that he thinks are crucial in getting over the hump from ontological objectivity to ontological subjectivity. He calls them "awareness" and "consciousness". But, as defined, neither has anything to do with awareness and consciousness as we are trying to explain them.

He seems to think that he is addressing the objection I just made when he imagines the objection that he does not explain why consciousness "feels" the way it feels. He thinks the problem is one of the *specificity* of this or that subjective state and he compares it to trying to explain "whether lions could have been made more gentle" or "eagles less hungry". However, that is not the problem. The problem is not why such and such feeling feels the way it does. The problem is, How can feelings exist at all? If we can answer that question, then the detailed specific questions will presumably receive detailed anatomical and physiological answers. The problem is not why red doesn't look blue, but how is consciousness possible at all, and within that question how is visual consciousness possible? If you had a complete answer to those questions, then you can address red and blue with detailed anatomical and physiological discussions of receptors, neurotransmitters, feedback mechanisms, etc.

Juraj Hvorecký: Causality and Free Will

Hvorecký raises far more questions in his paper than I can hope to discuss in this brief reply. I will reserve my comments for areas where I think we actually disagree. There are so many areas on which we agree that it is hardly worth commenting on them, so I will comment on only two of the many points he makes. First, he seems to think that panpsy-

chism is a well-defined notion. I do not think it is. In fact I think it is literally incoherent. The reason can be stated fairly simply. Consciousness always comes in organized conscious fields. There is a point where my consciousness ends and yours begins. And because consciousness always comes in units, the problem of panpsychism is that it cannot state what the unit is. So if you think that the thermostat might be conscious, then why not each screw in the thermostat? Or the whole heating system in which the thermostat is a part? Or the building that the heating system is in? Or the whole city? If you do think, "Yes each of these is conscious too", then what is the relationship between their consciousnesses and the consciousness of the thermostat? Consider the example of one's own body. My brain is clearly conscious, but what about my feet, my heart, my legs, my stomach? Are they conscious too? And if so, What is the relationship between each of their consciousnesses and the consciousness in my brain? The problem with panpsychism is not just that it is false but that it is incoherent. Consciousness comes in units and panpsychism cannot state what the units are.

The second point where I think we may disagree is on the notion of causation. He quotes Russell's famous claim that the word "cause" does not occur in advanced physics, but I think the fact that the word "cause" seldom occurs in advanced sciences is really irrelevant to the fact that causation is essential to physics and other natural sciences. It is true that when we are doing well in science we use mathematical equations. But often, indeed typically, those equations state causal relations. So for example Newton's inverse square law is a causal law. But it is not necessary to use the word "cause", because the notions in the law are causal to start with – bodies *attract* with a certain *force*, etc. When we speak of the four basic forces in the universe – the weak and strong nuclear forces, electromagnetism, and gravity – all of those are causal forces. The actual word "cause" is used more in practical areas of investigation, as when we look for the causes of cancer or the cause of AIDS; but the more theoretical domains, where we use generalizations and laws, contain the notion of cause just as much as do the practical domains. Perhaps a more serious disagreement comes over his understanding of my problem of the "gap". The word "gap" of course is a metaphor; I do not actually suppose that there could be holes in the brain corresponding to the causal gaps in the psychological formation of actions. The point is this: assume there are experiences of gaps in my sense, that is to say, experiences of making up your mind where you

sense that though you decided to do one thing you could have decided to do something else. Assume also that in these cases the causes of your deciding one thing were not causally sufficient to fix that particular decision or action. Assume also that in any given instant, the conscious processes are entirely caused by and realized in the lower level neurobiological processes. The word "supervenience" is perhaps unfortunate because it gives people the impression they understand something when generally they do not. But perhaps it is not too misleading to say that consciousness is supervenient on brain processes. Now, from these three assumptions, it follows that if the indeterminism at the psychological level is real – that is to say if the actions really are not caused by antecedently sufficient conditions – then there must be a corresponding absence of causally sufficient conditions at the lower level. If the neurobiology at any instant is sufficient to fix the psychology, and the psychological experience is "gappy"– that is, indeterministic – and if the experience of the gap is real – if it really corresponds to an absence of determinancy in nature – then there must be a corresponding absence of causally sufficient conditions at the lower level.

Another point of disagreement is that the fMRI scans that he takes as giving overwhelming evidence for Hypothesis 1 seems to me not at all conclusive, as he describes them. The fact that there is a time gap between increased neuronal activity and the agent's consciousness of a decision does not establish that the decision was fixed by causally sufficient conditions

Tomáš Hříbek: Thoughtful Brutes

Hříbek correctly describes the debate between me and Davidson over whether or not animals have thoughts. He is correct in thinking that it seems obvious to me that animals have thoughts and that I would regard it as proof of a philosophical error – a *reductio ad absurdum* of a theory – if you got the result that animals were incapable of thought. I once told Davidson that his view is not just bad philosophy, it is bad biology.

I also agree with Hříbek that Davidson's conception of what he calls "triangulation", as essential to cognition, is extremely implausible as an account of how humans think. It is not the case that all thought requires other people.

He thinks there is a problem for me in how we get from simple pre-linguistic forms of intentionality to complex linguistic forms of intentionality. I see this development as both a gradual process – from an evolutionary point of view and from a logical point of view – in that you can see how more complex forms are built on top of simpler forms. So, initially the animal has pre-linguistic forms of perception, intentional action, memory, belief, desire. They also have pre-linguistic forms of reasoning, such as reasoning how to achieve an objective by selecting the right means to achieve their ends. Köhler's experiments, as early as the First World War, are decisive in showing that animals have means-ends reasoning. Once an animal or tribe has linguistic forms of intentionality – even if they are in such simple forms as: "a man is approaching", "it is raining", "I am hungry", etc. – then it is easy to see how these could evolve into more complex forms. It is even easy to see how they could evolve into institutional forms. Such things as private property and marriage could evolve out of sheer physical possession and pair bonding. Indeed, it seems to me, once you have language it is pretty much inevitable that you will get institutional facts. Languages enable us to have all sorts of complex thoughts that we cannot have without language. But that is not inconsistent with the claim that the intentionality of language is itself based on pre-linguistic forms of intentionality. The relation of the complex to the simple is one of evolution. The complex forms evolved out of simpler forms. We do not know the details of how that in fact happens; but it does not seem at all philosophically difficult to suppose that it could happen, because we know as a matter of fact it did happen.

I have an additional terminological worry with his paper in that he accepts terminology that seems to me extremely dubious. The distinction between *de re* and *de dicto* is typically confused, and I have attacked it at some length (see Searle 1983, 208-217). There is a distinction between reports of beliefs that commit the reporter to the existence of an object that the belief is about and other reports that do not commit the reporter in this way. Consider the following:

1. About the girl next door, John believes she is nice.

and

2. John believes that the girl next door is nice.

These are two different reports of beliefs, but they do not mark two different kinds of belief. The belief is the same in both cases. The simplest proof that there cannot be distinction between the *de re* and the *de dicto* beliefs in these cases is that the person having the belief cannot make the distinction. He cannot say: "I have a *de re* belief: about the girl next door I believe she is nice, but I do not have a *de dicto* belief to the effect that the girl next door is nice." But to repeat, the distinction is not between kinds of beliefs but between kinds of reports of belief.

He assumes that there is a very precise meaning to whether or not something is propositional or not. An intentional state is probably best construed as propositional if it represents a whole state of affairs, but in that sense most intentional states are propositional, certainly perceptions and intentions-in-action. This usage, I think, probably runs counter to the way many philosophers think of propositions. So I think it is probably better for me, at least, not to use the notion of a proposition unless the context makes it absolutely clear what is at stake. I think his writing does not make it clear what he thinks is involved in having an intentional state with a propositional content. I do not think you can make sense out of current ethology, or for that matter out of Köhler's earlier experiments on apes, without assuming propositions in my sense. But I think lots of philosophers have a different sense.

Pavla Toráčová: Intentionality and What We Can Learn About It from Searle's Theory of Institutions

Toráčová begins her essay by pointing out that there is supposed to be a traditional problem about how intentionality is possible, and this problem arises from the fact that it seems mysterious that intentionality should exist at all. As she knows, I do not think it is mysterious. I think it is a biological phenomenon and should be regarded as such. It is no more mysterious than digestion. If you start your analysis of intentionality with very abstract beliefs, then intentionality must seem very mysterious. However, if you start with an animal feeling hungry or thirsty, then it does not seem so mysterious. Our intentional states are caused by processes in the brain, and they are realized in the brain as biological phenomena.

She points out that neither I nor anybody else has given an adequate account of how content determines conditions of satisfaction, and I think that is a crucial question. She thinks my theory of institutions

may help us with these questions. I think her summary of my views on intentionality are very accurate. She does an excellent job. One possible misunderstanding is that she seems to think that the only function of psychological mode is to determine direction of fit. However, there are other features to psychological mode than just the determination of direction of fit. For example, thinking that it is raining and seeing that it is raining have the same direction of fit, but the psychological mode determines different characteristics all together. Thinking something consciously and seeing are not the same thing. She is correct to emphasize that intentional content determines condition of satisfaction *under certain aspects* (p. 88). She is right to think that because institutional facts are created by intentionality, then "it is natural that to investigate the ontology of institutional facts amounts to an investigation of those beliefs, acceptances, recognitions, expectations, actions, etc…" (p. 89). She is also correct to see that institutional facts require 1) the imposition of function 2) the status character of the functions 3) collective character of the imposition and, I would add, 4) the acceptance. She asked the fascinating question, What is the character of the intentional states that imposes Status Functions? She says:

> The institutional facts have a special ontology because they are constructed as intentional objects of mental acts of some community. It seems that we can conclude that it is constructed in the performance of (at least) two intentional states of different kind: one of them is cognitive, the other is volitive. The institutional fact is constructed as their common product (p. 92)

I think she is onto something here. It was only in my later book, *Making the Social World,* that I saw what I now think is the right way to describe this is to say that they are invariably constructed by Status Function Declarations that have both directions of fit.

She concludes with an extremely provocative idea. I am not sure I understand it; I wish she would pursue it further. It says we relate to natural objects in two different ways: through a function relative to our goals and interests and through the properties we can perceive through our senses. Perhaps this is the essential condition under which a state can be intentional at all: to represent the same object in two different ways, and to do it in such a manner that it is exactly the interconnection of these two representations (i.e. the condition of satisfaction) that presents the object as identical. This is a fascinating idea, but it needs more thought. The direction in which I went is really quite different. I

suppose that every intentional state comes to us with a certain biological character, and then we evolve the capacity to perform Declarations that have both directions of fit; and these then enable us to construct institutional facts. She does not employ the concept of Declarations. She says, rather, there are two different ways in which we think about things; and this may be the essential condition in which a state can be intentional at all. I do not know what to say about this hypothesis. It is a very provocative and fascinating idea; I hope she will pursue it further.

Petr Koťátko: Searle's Defence of Internalism

I think Koťátko has written an excellent paper and I have no objections to it. I will confine these remarks to a few further thoughts.

It seems to me that the internalists are right in claiming that the internal contents of the mind are sufficient to fix reference and truth conditions generally; but in at least one sense of "content", it seems to me, the externalists are right about propositional content. Frege assumed that the sense of a whole sentence was identical with propositional content and that the sense of any referring expression determined the "mode of presentation of the referent". But mode of presentation and propositional content can come apart. If you consider the sentence, "I am here now", that sentence will express something analytic. On a normal usage it will express that the speaker is at the place and time of his spatially and temporally situated utterance. But the actual fact in the world that makes it the case that the speaker is at that place and at that time is not a necessary fact, and thus it seems that the analyticity of the utterance of the sentence does not carry over to the necessity of the fact represented. The fact is a contingent fact (I think this example was originally due to David Kaplan). So in at least one sense of content the externalists are right. However, there is another sense of content, where content provides a mode of presentation and where internalism is untouched by the externalist arguments.

He points out that there is Searlean style of responding to the externalists that says that these externally determined contents become part of the content of what is said, and that therefore the speaker himself is involved in the articulation of these contents. The right way to respond to the externalists is to say: "Many thanks for giving us a richer notion of the internal construal of content". I think that this is a correct conception, but I would add that there is a more specific flaw in the argu-

ments that I have seen for externalism. Most of these arguments fail to understand the precise character of indexicality in the determination of content. So there is a technical flaw in their argument. The central point, and I think Koťátko sees this, is whether or not all of these external features are determined internally.

One of the puzzling features of this whole discussion is how externalism became a contemporary orthodoxy, given that it is inherently implausible and the arguments for it are bad. Why is it so popular? I think the answer is that people thought correctly the traditional view of the meaning of general terms as given by a checklist of features – "water" is defined as colorless, tasteless, etc. – is inadequate. But the inadequacy of the checklist conception of meaning does not disprove an internalist account of meaning, and that, I think, is the mistake that led to the popularity of the externalist account.

III.

Zsofia Zvolenszky: Searle on Analyticity, Necessity and Proper Names

Zsofia Zvolenszky has produced an excellent paper and I am grateful for all the work and thought that went into it. My comments will be more than unusually inadequate, but my main comment is that I hope she will pursue these ideas further as I think she is making excellent progress.

She is right to say that when I wrote proper names in the mid 50's, I used necessity and analyticity as equivalent. This was very common at the time. She is quite right to point out that I did not use necessity in the sense of metaphysical necessity. She makes an interesting distinction between objects featuring truth conditions and descriptions featuring truth conditions. She argues that objects featuring truth conditions are closer to my purposes and that this approach avoids the commonly raised objections. I think this is a fascinating idea, and she is right to point out that for proper names it seems very plausible because of course one of the functions of a proper name is to pick out an object, not features of the object. I am very grateful for her contribution and I hope her ideas receive further attention.

I would add to it the following reflection which I think maybe significant. The entire subject of the philosophy of language since Frege,

and maybe since the Ancient Greeks, has suffered from a pervasive mistake. Before investigating issues in philosophy of language philosophers tend to assume that we are given a prior inventory of objects with their properties. Of course in a sense that is right. Before there was ever language there were objects and their properties. So given the assumption of an inventory of objects with their necessary and their contingent properties, the philosopher then asks, How do proper names name these objects? How do proper names refer to these objects? The reason this is objectionable is that the cognitive apparatus by which we are able to distinguish bits of reality into this or that object with this or that existence conditions or this or that identity conditions is part of the same cognitive apparatus by which we can refer to objects, name objects, and describe objects. In analyzing the nature of reference and description, we are not entitled to assume an inventory of objects. The reason is that the use of a name to refer to an object can only be explained if we answer the prior question, What is an object anyhow? The suggestion I am now making is that the answer to that question will go a long way toward answering the question, What is proper name of an object? Indeed, it is a necessary presupposition to answering the latter question.

Marek Nagy: The Role of Proper Names and Social Ontology

I do not have anything very useful to say about his fascinating paper. Analytic philosophers have treated names as just a special kind of referring device for identifying particular objects. Their obsession has been with such question as, How do they resemble and differ from definite descriptions and indexicals? He thinks that we ought to take more seriously the social and psychological significance of individual proper names of human beings and he also makes the striking suggestion that we ought to think of these as status functions. But a test for whether or not something is a Status Function is the deontic powers that attach in virtue of its having that status. The point is not just to that certain names such as "King George the 5th" are status indicators. That is plainly true. But I take it his suggestion is that the mere possession of a proper name is itself a kind of status. I think there may be something to this suggestion, but it would take more working out then either he or I has done. Also, his example suggests that cultures differ in the importance of a name. I have been told that in Japan there are a list of possible

surnames and that in the early days of Korean immigrants they found it difficult to get social services because they did not have a name on the list. This is clearly a case of a status function. Having one of the approved names confers a deontic power. But, the deontic power, I take it, derives from the fact that having one of the approved names constitutes being an authentic Japanese citizen. So the list of proper names is a list of status indicators.

Vít Gvoždiak: John Searle's Theory of Signs

I am grateful to Vít Gvoždiak for his attempt to relate my work to work in semiotics. I do not know enough about research in semiotics to have an intelligent opinion about it. Also there were substantial portions of his paper, for example the discussion of "Dynamical Objects", that I simply did not understand. My impression – but this is very much seen from outside – is that the semiotic apparatus is too crude to address the questions that interest me. The notion of a sign covers too many different types of things. He entitles his paper "John Searle's Theory of Sign" (I think he means signs, not sign), but I have no theory of signs. I have a theory of language. It is assimilating me too much to a semiotic paradigm to suppose that it is correctly described as a theory of signs.

If semiotics is to be concerned with "meaning" in the broadest sense, then it is not really accurate to say, as he does, that the distinguishing feature is that "semiotics is a study of every possible thing that can be used for lying" (p. 150). I think this is much too crude a criterion. Any type of human action at all *could be used for lying*, but that does not mean all human actions are semiotic. If I know you will think that when I wear a suit I am trying to appear to be rich, then I could "lie" by wearing a suit. Strictly speaking, lying occurs when I deliberately violate a *commitment* to truth. So if I say it is raining when I know it is not raining, that is a lie. There is a more general form of lying that covers just about any type of insincerity in commitment; so on this usage I could lie if I make a promise that I did not intend to keep. However it is a *reductio ad absurdum* to get the result that any act at all is semiotic, because any act at all could be used as a lie. This result is a *reductio* because we need to be able to distinguish the genuine cases of semantic content from those that have no semantic content. The important point about the possibility of lying is that it is a good test of certain classes of speech acts,

because it essentially requires the notion of a *commitment.* You can only lie if you can make a commitment that you can intentionally violate.

His example of the distinction between eye twitching caused by a neurological problem and winking as an intentional act is, I think, important; because the wink is a speech act. It is an intentional speech act, and it can have semantic content. The point I emphasize is that all speech acts have to be intentional. I gather he thinks that this is inconsistent with semiotics.

He is right in his conception that what I call Status Functions require collective intentionality: the creation of money, private property, or government is not something I can do all by myself. Also, I think he may not fully understand my distinction between semantic facts and other sorts of non-linguistic institutional facts. The sentence "Snow is white", can be used in virtue of its meaning to make the statement that snow is white. Analogously, the sentence "Obama is president", solely in virtue of its meaning can be used to state the fact that Obama is president. However, there is a huge difference between the fact that snow is white and the fact that Obama is president. The latter is created by representations, by semantics; but, the fact created goes beyond semantics in a way that the fact represented by "Snow is white" is just a fact in the world like any other. In the case of non-linguistic institutional facts we use semantics to create a reality that goes beyond semantics. So, the formula "X counts as Y in C" plays a completely different role in the creation of non-linguistic institutional facts – such as the fact that Obama is president – from the role it plays in purely linguistic institutional facts – such as the fact that the utterance of the sentence "Snow is white" counts as a making of a statement that snow is white. The terminology here makes this point an awkward thing to express, because all institutional facts are in a sense linguistic. However what I am trying to get across is that though they are linguistic in both their creation and maintenance; nonetheless the set of powers created go beyond linguistic or semantic powers.

Perhaps the best way I can respond to his thoughtful paper is to list certain principles that govern my research on language. The interesting question for this discussion then is: To what extent are they consistent with the semiotic approach?

1) All meaning is a matter of human intentionality. Meaningful symbols are always meaningful because meaning has been imposed on them by some conscious agent.

2) We need a crucial distinction between the *standing* conventional meaning of words, symbols, sentences, and the *utterance* meaning of a particular use of the symbol, word, or sentence to perform a speech act on a particular occasion.

3) A fascinating set of questions which I have attempted to answer, at least in part, concerns the systematic relations between *standing* conventional sentence or word meaning on the one hand, and speaker's *utterance* meaning on the other. How do we get from standing sentence meaning to speaker's utterance meaning when they differ? Cases where speaker's utterance meaning and standing conventional meaning come apart include: metaphor, indirect speech acts, and various other forms of figurative language. Utterance meaning also includes fiction, in a sense, because words in a fictional sentence can be uttered with their literal meaning; and yet the commitments normally carried by that literal meaning are absent. The ease with which people understand utterance meaning, even in cases where utterance meaning differs from sentence meaning, suggests that the relations between sentence meaning and utterance meaning are systematic. I have attempted in various writings to undertake an investigation of the systems in question, especially in *Expression and Meaning* (see Searle 1979).

4) The whole area of meaningful human acts, institutions, etc. that goes beyond semantics, in the strict sense, also seems to me a fascinating area of investigation; and in one sense, my whole research into social ontology is a matter of investigating meaningful social and institutional facts. The fact is created by the use of language, but the fact goes beyond language. Consider, for example, the fact that Obama is president, or that this is a five dollar bill. Both facts are created by language, but they are not linguistic facts.

Tomáš Koblížek: How to Make the Concepts Clear – Searle's Discussion with Derrida

I am especially grateful for Koblížek's discussion of my debate with Derrida, because unlike most of defenders of Derrida, he presents rational arguments in a clear and civilized fashion. I have not always found this either from Derrida or from his followers.

There are actually two points of disagreement between Koblížek and myself. I will briefly state and answer them. First, he defends Der-

rida's thesis that "Every concept that lays claim to any rigor whatso-
ever implies the alternative of 'all or nothing'" (p. 163).

To my objections that there are lots of rigorous concepts that allow
for imprecise boundaries and since Wittgenstein's *Philosophical Investi-
gations*, we have begun to develop interesting theories as to why that
is so; he maintains that Derrida's thesis is a thesis about the purity of
concepts and not about their *applicability*. The concepts are pure; it is
just the real world which is impure. He quotes Derrida saying that we
might give an analysis of promising that had the consequence that no
one in the history of the world had ever made a promise. On this view,
the concept of promising would be pure concept. Whether or not it ap-
plies to anything in reality is beside the point.

I think this is an extremely implausible conception of concepts, and
to any who advances it we would have to ask: Where does the con-
cept get its purity from? In fact, we understand concepts because we
understand how words are applied in actual cases. There are plenty
of rigorous concepts such as truth, metaphor, promising; and in each
case the concept is such that it applies more or less to particular cases.
In order to make his case seem plausible he would have to go through
some examples, and show how the purity of concepts is unsullied by
the imprecision of their application. I believe if you get the result that
no one in the history of the universe ever made a promise, you would
know that you had made a mistake, and you had better go back and
redo your analysis of promising.

I think that if you go through a number of examples you will real-
ize that the account he gives is incoherent. Here is why. The problem
is to account for marginal cases and the idea is that the marginality is a
feature of the real world but not a feature of the concepts themselves.
But because the whole point of the concept is to determine extensions it
would follow immediately – if one accepted the logic of "all of nothing"
– that any such concept has no marginal cases of extension, because the
logic of "all or nothing" has the consequence that the marginal cases are
all "nothing". The "all or nothing" character of the concepts is precisely
an "all or nothing" character of their application. So you cannot make
the kind of distinction between the concept and its application that he
supposes Derrida to be making. You would incidentally immediately
get inconsistencies. For example, both the concept of the literal and
non-literal are presumably "all or nothing" concepts, but since they

both admit of marginal cases, then it turns out these marginal cases are both literal and non-literal.

There is a very important philosophical point that needs to be made. Traditional metaphysicians, from Plato to Frege, thought of concepts as inhabiting an ideal realm unsullied by the sordidness of actual applications. We now think that, or at least I think, this is a total misconception. We should think of concepts, languages, etc., as part of our natural biological and social history; the lack of purity and precision is not a defect, it is an essential trait of their functioning. Much of Wittgenstein's *Philosophical Investigations* is devoted to rejecting the purity of concepts view, and I would be interested to see what response Koblížek and Derrida would make of Wittgenstein's arguments. As far as I can tell, Derrida never had any understanding of Wittgenstein.

In the second half of his paper he attempts to challenge the distinction between literal sentence meaning and speaker's utterance meaning. He thinks the distinction between sentence meaning and speaker meaning is the distinction between two different species of meaning, in the sense that, for example, cats and dogs are two different species of animal. However, that is not a correct conception of the distinction. The distinction is a category distinction between the *resources* provided by a language and the *use* of those resources on particular occasions. The language provides us with a finite stock of words and rules for combining them that enables us to generate an infinite number of sentences, and all of that is a matter of conventional word and sentence meaning. But the whole point of this is to enable people to communicate. People talk and write with these sentences, and that is where the question of what the speaker means comes into play. The distinction is a *category* distinction – like the distinction between an oar and the use of an oar to row a boat, or a tennis racket and the use of a tennis racket to play tennis.

If everybody always used sentences with only the precise and literal meaning of the sentence, then the distinction between sentence meaning and speaker meaning would be less useful to us. For example, it is not very useful in mathematics, though of course it applies in mathematics as a purely categorical point. Even in math there is a difference between the sentence "2+2=4" and actually claiming "2+2=4". However, in the actual operation of languages it is essential to see that speaker meaning often departs from the literal sentence meaning. The examples of these are quite famous: metaphor, indirect speech acts, other figures of speech, such as simile, metonymy, synecdoche, and a large number

of others. Indeed, one of the fascinating things in the philosophy of language is to work out the systematic relations between sentence meaning and speaker meaning. Notice that we typically have no difficulty communicating in cases where the speaker meaning differs from the sentence meaning, but that raises a theoretical question, How does it work? My second book on language, *Expression and Meaning* (see Searle 1979) was largely devoted to this question. One of the most fascinating cases is the case of fictional utterances where the words are used to mean what they normally mean, and yet the speaker's commitment is different from the normal commitment carried by the literal meaning of the sentences.

My conclusion, to summarize, is first that a theory of concepts has to allow for applications, more or less, *as part of the very structure of the concept*. Any theory of concepts that insist that all well-defined concepts must be "all or nothing" immediately has absurd results. Secondly, he is mistaken in supposing that sentence meaning and speaker meaning are two different kinds of meaning. The distinction is between the resources provided by language and the use of those resources in actual communication. The main purpose of language is to enable people to use sentences in communication, thought, etc. The purpose of having sentence meanings is to enable speaker meaning.

IV.

Jiří Koten: Searle's Approach to Fiction: Extending the Concept to Other Media

Jiří Koten extends my analysis of fiction to various other genres, particularly to film. I argued that in a work of fiction the author pretends to perform assertions and various other sorts of speech acts. Extending the analysis to dramatic productions, I argued that in a play the burden of pretense is borne by actors, and that the text of the play is best construed as a set of directions as to how the play is to be performed. The text of the play is, so to speak, a Directive, a recipe for performing a collective act of pretense by the actors. The question is: How does this extend to the cinema? It would be interesting to work this out, and I have not done so. The question is: Who bears the burden of pretense? Can I extend the analysis of theatrical productions to cinema? I think it does extend. As in a play on stage, the actors are engaged in a pretense; and

the author of the screen play and the director of the movie give them instructions as to how to carry out the pretense. I do not see any additional burden of pretending borne by them beyond that of the actors.

In *Casablanca*, for example, Humphrey Bogart and Ingrid Bergman pretend to be two unhappy lovers in war time. The makers of the movie simply record this and other events on film and distribute it. I do not see any additional pretense. There is an additional feature to film, and, that is, typically there is no single intelligence behind the production of the film. Even the most powerful directors can have their creative work altered or vetoed by the producers. The crucial question to ask is, Who is committed to what exactly? And I cannot see that any additional commitment of pretense is borne by the team that made the film or the company that owns the film.

In the philosophical analysis of fiction there is a remaining question that I have not addressed, and I would like to use this occasion to discuss it further: Why does fiction matter so much to us? We know the stories are not true. Koten says correctly that reading fiction is a lot of fun. It is very entertaining to read fictional stories. And that is right, but that does not account for the enormous importance attached to literature in our culture. In my case, in my own life, I would be a different sort of person altogether if I had never read Tolstoy, Dostoevsky, Hemingway, Faulkner, Joyce, Mann, Proust, Kafka, etc., etc. I do not believe there is any single answer to that question, but at least part of the answer is this: reading works of fiction we become imaginatively engaged with fictional characters, and we acquire about them an epistemic intimacy that is difficult, if not impossible, for characters in real life. The paradox is this: precisely because the person is imaginatively created, and we share in the imaginative creation of the author, we get a kind of closeness and intimacy to the fictional characters that we cannot get, except to a few people, in our real lives. I feel I know Emma Bovary, Hans Castorp, Holden Caulfield, and Jay Gatsby better than I know many members of the Berkeley Philosophy Department. Also, because the Background in a work of fiction can be so different from anything the reader lives in that by imaginatively becoming involved in the lives of the characters the reader expands his own sensibility. Life in Paris over one hundred years ago, as described by Marcel, is quite different from any environment I had grown up in when I first read Proust. But I became completely familiar with and at home in the milieu.

The philosophical problem is this: if none of it is true, why does it matter so much? Here is part of the answer: precisely because the character is fictional, precisely because he or she is imaginatively created, we can get closer to them than we can to most people in real life. The events of real life come to us uninterpreted, but in a work of fiction we are already presented with the author's interpretation of the events that he describes. As a result the inexhaustible ambiguity of real life is seriously circumscribed. We can still add further interpretations to the author's creation. We are much closer than we are in the maelstrom of our real social existence.

It is an important fact about fiction, unlike poetry, that just about all of it is about people and their lives. Plenty of poems are entirely about the starry nights, the landscape, the sea, etc.; but virtually no works of prose fiction are about starry nights, the landscape, the sea, etc. They are about people and the events of their lives, and I believe they matter so much us because we get so close to them. It might seem then that I am suggesting there is a kind of voyeurism in our interest in fiction, and that again seems to me exactly wrong. It is precisely because we know that the characters we care so much about are not real that we know perfectly well that we have no obligations whatever to respect their privacy, to prevent ourselves from learning any of their secrets. Not only are we closer to many people in fiction than we are to real people, but we are closer in a way that avoids any moral obligation to them whatever. The reader pondering the characters in a novel is not at all like the peeping Tom looking in at the window. The peeping Tom knows very well that he is part of the total scene and that his behavior is reprehensible. The reader of the novel knows very well that he is in no sense part of the scene in the novel, and that he is not in any way under a moral obligation to respect the privacy of the characters in the novel. On the contrary, the whole point of reading the novel is so that he can become imaginatively connected to the lives of the characters. This is not the only reason that the great works of fiction matter so much to us, but I am convinced that it is one reason.

Jan Tlustý: Fictional and Factual Autobiography from the Perspective of Speech Act Theory

The paper of Jan Tlustý is fascinating. I have not read all of the works that he refers to so I cannot really comment on them. The general

question he is addressing concerns the relation of factual and fictional autobiography. The general theory that follows from my conception of speech acts is that the crucial question is, To what is the author committed? Or in Tlusty's phrase, For what is he held responsible? That is not at all an obvious or simple question in many cases, so let us consider a book that I am familiar with, Coetzee's book *Summertime.*

According to Genette in the case of factual autobiography the author is identical with the narrator who is identical with the chief character A = N = C. Utterances in such works are fully committed assertions. However, if the author chooses to be writing in the third person, then he ceases to be identical with the narrator and the work is now a standard work of fiction, even if it is autobiographical in content.

But, in the case of Coetzee, the situation is complicated by the fact that though the author is not identical with the narrator and is identical with the chief character, the narrator is himself a second author who is writing a biography of the author. That is, Coetzee is really writing a work of fiction and in that work of fiction the narrator Vincent is writing a biography of the author Coetzee. The situation is further complicated by the fact that the research method of the biographer is to interview people who were close to Coetzee and learn the details of his life. I think in this case we cannot treat the text as just another work of fiction. It is unlike Coetzee's novel *The Diary of a Bad Year*, which is genuinely an autobiographical novel in the form of a work of fiction.

In *Summertime*, I think the author assumes responsibility for general conception of the life of the main character of the fictional biography written by Vincent. Various clues tell us that the work is autobiographical. Indeed, the clues are not very subtle: the name of the author and the name of the subject of the biography are the same, J.M. Coetzee. The details of their life are pretty much the same. I think we would hold the author responsible for distortion if we discovered there were systematic distortions. We would feel this was a flaw in a way that we would not feel if he if the character in *Diary of a Bad Year* was not really like the real Coetzee

I think that Tlustý's understanding of this work is very profound. When he interprets Coetzee saying:

> … grasping one's life, as well as the life of others, is complicated, and fraught with peril, misunderstanding and misinterpretation. Is the story of our life the story that *we* tell about ourselves, or is it the story that *others* tell about us? (p. 184)

I agree with him about the powerful aesthetic effect of the work. There are three books that are all part of the same autobiography. I am not sure that he is right in thinking that the serialization of the three books under the one subtitle *Scenes from Provincial Life: Boyhood, Youth, Summertime* accentuates the fictionalization; it seems to, if anything, emphasize the autobiographical character.

Jakub Mácha: Searle on Metaphor

Mácha gives a generally accurate statement of the distinction between me and Davidson on the subject of metaphor. Essential to my account is that metaphorical utterances are cases of speaker meaning not sentence meaning. When Davidson says, "metaphors mean what words literally mean," I think he is right about sentence and word meaning but not right about speaker's utterance meaning. It used to be commonly said in the literature on metaphor that in a metaphorical utterance at least one word changes its meaning; but that is not true, because if it did change its meaning, it would no longer be a metaphor. The whole notion of metaphor is the notion of using a word that has one conventional word meaning with a different speaker meaning in that metaphorical utterance. The dispute between me and Davidson is that he is committed to denying that there are metaphorical speaker meanings of utterances. That seems to me plainly mistaken.

As I use these expressions, Mácha is mistaken to say that metaphorical utterances are indirect speech acts. A typical indirect speech act, such as "Can you shut the door?" or "Can you pass the salt?", is one where the speaker means what he says but means something more. In indirect speech acts, speaker meaning *includes* literal sentence meaning but goes beyond it, whereas in metaphors, the typical metaphorical utterance the speaker *does not* mean literally what he says but has a metaphorical utterance meaning. Of course an utterance can be both a metaphor and an indirect speech act, but the distinction should still be clear.

It is quite right to say that the "rules" for interpreting metaphor are very unstrict. It is best not to think of them as rules at all, but as sets of procedures by which people can recognize the speaker's metaphorical utterance meaning given the fact that the sentence was not uttered literally.

His account of my reply to Davidson's rejection of metaphorical meaning is, I think, exactly right: the concept of metaphorical meaning has explanatory power. The decisive objection to Davidson, as a general account of metaphor, is that metaphorical utterances can be semantically evaluated as true or false. So if I say "Sam is a pig" or "Sally is a block of ice", the metaphorical content of these utterances may be debated and agreed with or disagreed with; a conclusion as to truth or falsity may be reached even though the utterances are metaphorical. It is true that there will always be certain open-endedness to metaphorical speaker's meaning but that is characteristic of literal utterances as well.

In general, I am sympathetic to his whole account, I thought it was careful and perceptive, I hope he will pursue these matters further.

References

SEARLE, J. R. (1983): *Intentionality: An Essay in the Philosophy of Mind.* Cambridge: Cambridge University Press.

SEARLE, J. R. (1979): *Expression and Meaning: Studies in the Theory of Speech Acts.* Cambridge: Cambridge University Press.