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.

Organon F 19 (2012), 199-225

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 disagreement is not genuine disagreement. I think in the end we probably agree, and I have not succeeded in making my views sufficiently clear.

Vladimir Havlík: Searle On Emergence

Havlík discusses my conception of emergence. I draw various analogies. 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 reducible, 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 notion 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 property 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 relationship 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 whatsoever 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 Investigations*, 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 applies 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 concept 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 realize 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.