Эпистемология и философия науки
2019. Т. 56. № 3. С. 138–152
УДК 161.1

Epistemology & Philosophy of Science
2019, vol. 56, no. 3, pp. 138–152
DOI:

COMPUTATIONAL KNOWLEDGE REPRESENTATION IN COGNITIVE SCIENCE

Igor F. Mikhailov –
CSc in Philosophy,
senior research fellow.
Institute of Philosophy,
Russian Academy of Sciences.
12/1 Goncharnaya St.,
Moscow, 109240,
Russian Federation;
e-mail: ifmikhailov@gmail.com

Cognitive research can contribute to the formal epistemological study of knowledge representation inasmuch as, firstly, it may be regarded as a descriptive science of the very same subject as that, of which formal epistemology is a normative one. And, secondly, the notion of representation plays a constitutive role in both disciplines, though differing therein in shades of its meaning. Representation, in my view, makes sense only being paired with computation. A process may be viewed as computational if it adheres to some algorithm and is substrate-independent. Traditionally, psychology is not directly determined by neuroscience, sticking to functional or dynamical analyses in the what-level and skipping mechanistic explanations in the how-level. Therefore, any version of computational approach in psychology is a very promising move in connecting the two scientific realms. On the other hand, the digital and linear computational approach of the classical cognitive science is of little help in this way, as it is not biologically realistic. Thus, what is needed there on the methodological level, is a shift from classical Turing-style computationalism to a generic computational theory that would comprehend the complicated architecture of neuronal computations. To this end, the cutting-edge cognitive neuroscience is in need of a satisfactory mathematical theory applicable to natural, particularly neuronal, computations. Computational systems may be construed as natural or artificial devices that use some physical processes on their lower levels as atomic operations for algorithmic processes on their higher levels. A cognitive system is a multi-level mechanism, in which linguistic, visual and other processors are built on numerous levels of more elementary operations, which ultimately boil down to atomic neural spikes. The hypothesis defended in this paper is that knowledge derives not only from an individual computational device, such as a brain, but also from the social communication system that, in its turn, may be presented as a kind of supercomputer of the parallel network architecture. Therefore, a plausible account of knowledge production and exchange must base on some mathematical theory of social computations, along with that of natural, particularly neuronal, ones.

Keywords: computation, knowledge, cognitive science, formal epistemology, neuroscience, semantics

ВЫЧИСЛИТЕЛЬНАЯ РЕПРЕЗЕНТАЦИЯ ЗНАНИЙ В КОГНИТИВНОЙ НАУКЕ

Михайлов Игорь Феликсович – кандидат философских наук, старший научный сотрудник.
Институт философии РАН.

Когнитивные исследования могут внести вклад в формально-эпистемологическое исследование репрезентации знаний, поскольку, во-первых, они могут рассматриваться как описательная наука, имеющая тот же предмет, для которого формальная эпистемология является наукой нормативной. И, во-вторых,

Российская Федерация,
109240, г. Москва,
ул. Гончарная, д. 12, стр. 1;
e-mail: ifmikhailov@gmail.com

понятие репрезентации играет определяющую роль в обеих дисциплинах, хотя и различается в оттенках своего значения. Репрезентация, на мой взгляд, имеет смысл только в сочетании с вычислениями. Процесс может рассматриваться как вычислительный, если он придерживается некоторого алгоритма и не зависит от материального субстрата. Традиционно психология не определяется непосредственно нейробиологией, оставаясь на уровне функционального или динамического анализа на уровне «что» и пропуская механистические объяснения на уровне «как». Поэтому любая версия вычислительного подхода в психологии является весьма перспективным шагом в интеграции двух научных сфер. С другой стороны, цифровой и линейный вычислительный подход классической когнитивной науки мало чем может помочь в этом, поскольку ему не достает биологической реалистичности. Таким образом, на методологическом уровне необходим переход от классического вычислительного подхода в стиле Тьюринга к общей вычислительной теории, которая смогла бы охватить сложную архитектуру нейронных вычислений. Для достижения этой цели передовая когнитивная нейробиология нуждается в удовлетворительной математической теории, применимой к естественным, особенно нейронным, вычислениям. Вычислительные системы могут быть истолкованы как естественные или искусственные устройства, которые используют некоторые физические процессы на своих более низких уровнях в качестве атомарных операций для алгоритмических процессов на своих более высоких уровнях. Когнитивная система – это многоуровневый механизм, в котором лингвистические, визуальные и другие процессоры надстраиваются над многочисленными уровнями более элементарных операций, которые в конечном итоге сводятся к элементарным нейронным спайкам. Гипотеза, отстаиваемая в этой статье, состоит в том, что знания порождаются не только отдельным вычислительным устройством, например мозгом, но также и системой социальной коммуникации, которая, в свою очередь, может быть представлена в качестве своего рода суперкомпьютера, имеющего архитектуру параллельной сети. Следовательно, правдоподобное описание производства и обмена знаниями должно опираться на некоторую математическую теорию социальных вычислений, а также теорию естественных, особенно нейронных, вычислений.

Ключевые слова: вычисления, знания, когнитивная наука, формальная эпистемология, нейробиология, семантика

Representation of knowledge is an important problem from both metaphilosophical and methodological points of view. Its metaphilosophical importance stems from the fact that philosophy, together with its constituents and offsprings (*e.g.*, epistemology, formal epistemology, epistemic logical calculi, etc.), strives for conceptual solutions of what knowledge is or should be. The importance of this problem for philosophy of science is multifaceted, but in particular, I would highlight the fact that, upon its transition from the philosophical to the scientific level, the concept of knowledge often loses its modality that distinguishes it from that of belief: one can *believe* that 2*2=5, but one cannot *know* it.

As a result, in cognitive or computer sciences, when they speak about knowledge representation, they often mean just representation of information or propositional content.

Besides the question of what is represented in particular (*i.e.*, what *knowledge* is), we should also ask what it is to be *represented*. In this paper, I primarily concentrate on this matter. I will try to show that, of all the possible concepts of the relation between representation and representata – which may be causal, semantical, or computational – only the latter is relevant and productive for philosophical or scientific research.

A special attention will be drawn hereafter to the topic of research levels, by which I mean philosophy (in the broad and complex sense), cognitive science[1] (CS) and neuroscience (NS). In the actual state of the named disciplines, there is scarcely any trans-level integrity in understanding some basic notions, such as representations. There are philosophical discussions on (anti)representationalism in, e.g., theories of qualia, and there is a much more technical approach to representations in CS, also met with various 'radical' dismissals thereof [Hutto, 2011; Hutto & Myin, 2013] and we have an emerging field of computational NS where 'computations' and 'representations' have eventually obtained resident permit [Piccinini, 2018]. The problem is that the concepts labelled with the same term may differ essentially on different levels. And speaking about representation of knowledge in particular, it is no surprise that, going top down through the levels, this concept loses its philosophical glamour by turning into a technical idea of storing and representing some propositional (descriptive) content and some procedural mastery [*more on this see* Kovalyov & Rodin, 2018]. I will consider some philosophical issues with conceptualising knowledge and its formal explications with the view of its usage in computer-driven practices, and then I will propose some approach to saving epistemic and epistemological definitions of knowledge representations on the cognitive and neural levels.

Epistemology and Formal Representations of Knowledge

Our top-down discourse on knowledge representations implies that a connection between the philosophical level of epistemology and the scientific level of CS must be identified. In my view, one point of this connection is most probably what is usually referred to as *formal epistemology (FE)*. The reason for this is that CS as it is widely construed is based on the so called 'computer metaphor', which consists in the explicitly or implicitly

[1] This term refers here and after to methodological assumptions that underlie modern-style cognitive psychology, cognitive linguistics and Artificial Intelligence (AI).

adopted belief that the human or animal mind is structurally analogous to computing machines of a certain architecture[2]. On the other hand, there is a well-defined research field aimed at operationalising the concept of ‘knowledge’ for its further usage in knowledge-based systems of the AI realm – the science of computer knowledge representation. Therefore, ideally, there must be ontological and methodological intersections of CS and the knowledge-based AI, though, in fact, it is not always the case. But, anyway, if anything in the realm of philosophy can be of use for this kind of research, it certainly must be related to formal explications of knowledge ready to be translated into computational algorithms. FE is the best candidate thereto, as it makes quite a productive use of formal logical and mathematical techniques to achieve its goals – the same tools being used in computer knowledge-representation science.

In fact, most of FE research revolves around the modal aspect of knowledge mentioned above – the one that distinguishes it from a belief or an opinion. As far as I can tell, there are two principal approaches to identifying and explicating this aspect: the one that concentrates on logical and semantical properties of what may virtually be called ‘knowledge,’ and the other one that takes the communicative context of ‘knowing’ and ‘ignoring’ into consideration. The first one sticks to logical and formal semantical techniques and arguments, while the second one uses multi-agent epistemic logic or the game theory to conceive the functioning of knowledge in communicative contexts. For the adepts of the first one, knowledge is a product of a reasoning mind, while for the followers of the second one, it stems from agents’ mutual dispositions in their intercourse.

For the sake of the argument, these two positions may be reduced to the following disjunction: knowledge is either a property of a knower, or a relation between knowers. Let us briefly examine each one of the disjuncts.

Within the first conception, all that is needed for there to be knowledge are a knower and what is to be known. A knower, marked conventionally as *S*, is in a certain relation to a certain case marked conventionally as *p*. *S* may know that *p*, doubt that *p*, assume that *p*, hope for *p* to be (or not to be) the case, etc. In all of these occurrences, *S* possesses what may be called an *account* of *p*, plus some *attitude* to it. It is widely accepted that the account is invariant in all the occurrences, while attitudes vary. We may easily determine that what distinguishes the *knowing* attitude from the others is the fact that *S* cannot stand in this attitude to *p*, if *p* is not the case – unlike the attitudes of doubt, assumption, aspiration, fear, etc. This makes us believe that all that accounts for knowledge is the account of *p* added with some *sufficient evidence* for *p* to be true, thus

[2] The question of which architecture it is – *e.g.,* serial or parallel – is one of the main debates in CS.

making up a conception famously identified as ‘justified true belief’ (JTB) that allegedly goes back to Plato. The widely discussed Gettier (counter)arguments [*for discussion see, e.g.*, Veber, 2014; Henderson and Horgan, 2011, pp. 38–42; Nikiforov, 2018] show that there may be instances of JTB that all our intuition refuses to recognize as those of knowing. But a more fundamental flaw seems to be characteristic of the approach under discussion. This is the mystifying nature of the *sufficient evidence*. What measure of evidence is enough for a fact to be *known*, as opposed to *just believed*, *assumed* or *doubted*? If we say that a witness *knows* who is the murderer even because of having observed the act of murder, what in his/her position crucially distinguishes ‘knows’ from ‘strongly believes’ or ‘is absolutely certain’? Remember the *‘12 Angry Men’* classic.

Moreover, this ‘sole knower’ conception cannot generally deal with the fact of what I would call a *grammatical asymmetry of ignorance*. Suppose we transform ‘*S* knows that *p*’ into a conjunction ‘*p* is the case, and *S* knows that’. It is not a problem to change this sentence into the first grammatical person: ‘*p* is the case, and *I* know that’. But suppose our *S* is ignorant of *p*. It is no less legitimate to say ‘*p* is the case, but *S* doesn’t know that’. But the already familiar first-person transformation now leads to the nonsensical ‘*p* is the case, but *I* don’t know that’. This asymmetry takes place even if both *S* and *I* have or miss the same evidence for believing that *p* is the case.

So, we may come to a conclusion that the ‘sole knower’ approach misses something essential about knowing. The FE studies charged with this concept usually concentrate on the issues of knowledge ‘justification’ implying thereby various techniques of semantic or syntactic inference. If we assume that FE must provide some output for the research in computer knowledge representation, then all we can derive from the approach under discussion is the claim for propositions to be stored and processed by computer or AI systems together with inference procedures that would certify them as units of knowledge. It is quite a challenge to think of any pragmatical contexts where such an addition would make a system more adaptive or efficient. This may be one of the reasons why computer science generally omits all the philosophical buzz about knowledge treating it just as pieces of information to be stored and processed.

The other approach may roughly be called ‘multi-agent,’ as most of the FE arguments in its favour I borrow from the literature on multi-agent systems (MAS). Though this research field rather belongs to the second, properly scientific level of our analysis, it displays high demand of epistemological foundations and, consequently, great attention to the issues of epistemic logic, Kripke or possible world semantics, game theory and other FE tools. So, its implications related to the subject-matter of knowledge representation may well be promoted to the philosophical level of our discussion.

The extensive involvement of modal and epistemic logics in the literature on MAS is intended to cope with problems of common and distributed knowledge that are critical for systems of this kind[3]. Michael Wooldridge, the author of a comprehensive guide on MAS, proposes a formalism, which is a traditional first-order propositional logic with the addition of a set of unary modal operators K_i that read as ‘agent *i* knows that …’ [Wooldridge, 2002, p. 279]. He further shows that, when interpreted on the MAS, this logic encounters some semantic problems. For example, within unreliable communication, the epistemic state known as ‘common knowledge’ – ‘everyone knows that *p*, and everyone knows that everyone knows it’ – may turn out to be unaccessible being lost in endless iterations of mutual confirmations. Similarly, distributed knowledge may turn out to be problematic, when agent *i* knows that *A* and agent *k* knows that *A->B*. In this case, the system contains a knowledge that *B* in the distributed form. To formalise this knowledge, Wooldridge offers a special epistemic operator *D,* whose semantic definition involves not a set union, as in traditional modal logics, but intersection of epistemic worlds w_i of all agents in the system. In his opinion, “a *restriction* on possible worlds generally means an *increase* in knowledge” [Wooldridge, 2002, p. 283]. No less important, from his point of view, is the fact that the semantics of possible worlds, which is mainly used to interpret epistemic and modal logics, implies a reasoning agent being a ‘perfect logician’ who, *e.g.*, will see a contradiction of *A* and *~B,* if it is known that *A->B*. However, obviously, real agents, including the majority of living people, are often quite tolerant of such implicit contradictions, which does not prevent them from functioning. In view of that, Wooldridge proposes to restrict the MAS standards to the requirement of *weak consistency,* which would prohibit only the apparent contradictions of *A* and *~A* [Wooldridge, 2002, p. 276].

Another interesting formalisation is proposed in [Vlassis 2007]. Analysing the concept of common knowledge as exemplified in a well-known logical problem about three players in hats, each of whom sees only the colour of the hats of the two others, Nikos Vlassis introduces the following definition. Let *S* be the set of all possible states in general, of which only *s* is the actual state. Let also *i* be the ordinal number of an agent from some finite set. Each agent *i* sees the state *s* through an ‘information function’ $P_i(s)$ that generates a subset of *S* that includes only the states considered possible by the agent having only limited information available. Let also *E* be some subset of *S*, which we will call an *event*. K_i is a ‘knowledge’ operator assigned to a specific agent *i*. Then, by definition:

$$K_i\,(E) = \{s \in S\colon P_i(s) \subseteq E\} \text{ [Vlassis, 2007, p. 39],}$$

[3] By the way, the problem of omniscience that is of theoretical importance in FE [Fagin et al., 1995, pp. 333–390] also stimulates search for practical computational solutions in the MAS realm.

i.e., agent *i* knows *E*, if its information function P_i in the true state *s* contains *E*. Or, in the natural language, *some agent knows some event, if the set of all states seeming possible from its perspective is contained in this event*.

According to Vlassis, the definition proposed corresponds to that used in epistemic logic. The latter assumes that an agent knows the fact of φ, if φ is the case in all states that the agent considers possible. In an event-based approach, an agent knows an event *E*, if all the states that the agent considers possible are contained in *E*. Vlassis refers to the foundational work [Fagin et al., 1995][4], where it is shown that both approaches – logical and eventual – are equivalent [Vlassis, 2007, p. 39].

The application of the described formalism to the problem of players in hats provides a nice illustration to the principle of the *inverse relation of knowledge and possible worlds*, formulated by Wooldridge: *each added knowledge reduces the number of possible states in the agent's perspective*. However, agents' limited rationality relates not only to the amount of knowledge an agent has, but also to the quality of its reasoning. In the Vlassis model, all players in hats are 'perfect logicians' in Wooldridge's terms: they masterfully and consistently draw conclusions by applying the laws of non-contradiction and the excluded middle. At the same time, it is obvious that real agents whose behaviour is modelled in the MAS are rarely so. Therefore, the theoretical foundations of the distributed AI, one of which being FE, include fuzzy logic, statistical theories, and some other theoretical tools that allow us to bring the model closer to the complex reality.

The multi-agent approach, in my brief sketch, has its own flaws. In particular, it actually lacks distinction of knowledge and belief in the scope of a single agent, which counters our everyday intuition. But, at the same time, it offers an effective way of defining 'knowing' modality as a relation of an agent's epistemic world to the common or distributed knowledge of the system. In simple terms, one *believes* that *p*, if *p* is part of all possible states as seen from his/her perspective. But if *p*, besides this, is also contained in some way in the shared knowledge of the system, we may say that the agent *knows* it. One may object that this interpretation keeps the JTB frame, but just substitutes the Correspondence theory of truth with the Coherence one. The reply might be that all this argument is not so much about the truth, but rather about *social certification* of what is adopted as knowledge. And a particular mode of such certification includes not only particular truth conditions, but also particular ways of justification. Such an approach is perfectly aligned with the difference of ways and habits in acquiring knowledge as implemented in different cultures.

[4] Wooldridge also refers to this book.

Representations in Cognitive Science

The epistemological part of this study makes for the conceptual grounds of knowledge representation. I further assume that a review of cognitive discussion of representation as such forms a modelling frame, while that of neural research provides empirical evidence thereof. In the cognitive part I will focus on the issues of historically primary cognitive paradigm, usually referred to as *classicism*, as opposed to another one known as *connectionism*. They are not the only contestants on the cognitive field, but I consider them foundational as they appeal to the two possible architectures of computation and representation: the serial or the parallel computations and, respectively, joint or distributed representations.

The principal problem with the classicist CS stems from the historical fact that the widely discussed 'computer metaphor' of the human mind was preceded by 'human metaphor(s)' of the computer, which made the historically first computational architecture seem a suitable explaining model for the psyche. Other cognitive paradigms apply or base on other computational theories, such as that of parallel digital, parallel analog, statistical and other computations.

In the classical CS, as Nir Fresco points out [Fresco, 2012, p. 356], representations must have two important properties – to be *physically realisable* and to be *intentional*. Intentionality is also understood in a classical way – as the presence of meaning or content, that is, the representation of what it is. Physical realisability presupposes the presence of physically acceptable vehicles of representations, which may be computational structures or states of the brain. Within this view, representations, indeed, are physically embodied entities with semantics – that is, *symbols*. Turing-computable operations may be executed over them, and the entire model of cognitive acts is completely analogous to the work of a von Neumann computer. The obvious advantages of the classical model include its direct computer realisability: one of the founders of cognitive classicism, J.R. Anderson created the ACT-R computer platform for modelling cognitive functions to the end of subsequent experimental verification of models [Anderson, 1983]. Its explanatory principle is based on scientific abduction: if the model shows the same results as the living subject, then, with a high degree of probability, cognitive devices of the subject has the same structure as the computer model.

And here is where problems begin. Historically, the cognitivist paradigm triumphs after the victorious debate of N. Chomsky against B.F. Skinner in the late 1950s. The theory of innate generative grammar allegedly explained the productivity of human language – *i.e.*, its ability to compose and understand previously unheard statements. Linguistics defeated neo-behaviorist psychology in alliance with the rising computer science. Not surprisingly, the new cognitive approach had pronounced linguistic ancestral features: e.g., construal of mental life as a flow of computational operations on semantically loaded symbols. A legitimate consequence of this view was the concept of the 'language of thought' (mostly referred to as *LoT*, or *Mentalese*) by Jerry Fodor, according to which our operations with external symbols correspond to intrinsic manipulations with symbolic representations, such that these representations are obviously semiotic, and the operations with them are akin to statements in the natural language. *LoT*, according to Fodor, is the basis of our understanding of the language of communication [Fodor, 1975; 2008]. Such a position should not necessarily lead to bad infinity, especially since Fred Atteneave in 1959 presented a mechanistic model of a cognitive device that allowed to avoid the *homunculus paradox* due to the redistribution of functions between organisational levels of the system [Attneave, 1961]. However, despite overcoming the paradox at the design level, it is still threatening at the conceptual level, being formulated as follows:

> *(HP1)* To recognise an external content behind a symbol, one needs to have cognitive capacities. But they are precisely what we try to explain with this very scheme.

Suppose we can find a technical explanation of how the cognitive system recognises syntactic properties of intrinsic symbols. But where does the content that makes them intentional come from? And who reads this content? In short, there is a serious suspicion that the explanandum is contained in the explanans. As Fresco notes, "extrinsic representations are external-knower-dependent: a knower assigns external (or real-world) semantics to data structures, strings or symbols." (Fresco, 2012, p. 358). It is therefore no accident that research within the symbolic (classicist) paradigm has most succeeded in explaining linguistic capacities and linguistic activity.

Moreover, the ambiguity of the very notion of representation remains unsurmounted: is the state of a cognitive device or a mental (phenomenal) state meant thereby, or, in other words, do we speak of *objective* or *subjective* representations? The latter appear to be a more legitimate area of representations, since they usually stand in for objective states of affairs in a subject's mental vision (although this is not always the case either). As for objectively recorded states of cognitive devices, in my opinion, this view of representation plays a normative role in classicism: everyone assumes that such representations should be there, as they are provisioned in computer models applied.

Some of brain reading projects provide empirical results demonstrating functional relations between activation patterns of certain brain regions and external stimuli. Thus, in [Pasley et al., 2012] an attempt was made to demonstrate, via mathematical modelling, such relations between the spoken word and the activation pattern of the upper temporal gyrus, responsible for high-level processing of semantically laden acoustic information. Patients who were undergoing brain surgery because of epilepsy or brain tumors had sensors implanted into this area, with which one could reconstruct the structure of neuronal activations that arise when the patient hears real or made-up words. Next, a few mathematical models were built describing the functional relations between activation patterns and waveforms of spoken words. Then a case-relevant model was used to reverse the reconstruction of an acoustic image from neural pulses. The result was ambiguous: recovered sound forms of words went mainly unrecognised by listeners, but visually, however, pictured waveforms of recovered words were seen as corresponding to those of words actually uttered. The researchers suggested that, with the improvement of technical and mathematical tools, one will develop tools of communication with patients speechless due, for example, to paralyses.

Obviously, for a particular scientific field this result is intermediate. But in a conceptual analysis, we may assume that the empirical search has been successful, and a method of translating both waveforms of words into neural ensembles and vice versa is found. Then we must accept that the structure of activated neurons ensemble is, in the strict sense, an objectively recorded representation of the sound of a spoken word. And this, most likely, will be fair. But a so construed ‘representation’ is not a sufficiently operationalised concept for cognitive research and does not provide sufficient conceptual tools for solving philosophical and cognitive psychological problems related to mind and its numerous riddles.

There are several reasons for this. First, in this case, a neural activation pattern is just as much a representation of a sound of a spoken word, as, on the contrary, sound vibrations are a representation of an activation pattern of a neural ensemble. And, with this consideration alone, there is nothing specifically cognitive in the very concept of representation. Secondly, such an expanded, or ‘weak’, understanding of representation leads to pan-representationalism, as an analogy to pan-computationalism. Brain structures can be considered representations of external events on the same grounds as a synthesised protein can be considered a representation of a chain of RNA nucleotides – or vice versa, which is not important. Thus, a concept covering a wide spectrum of non-cognitive phenomena is put in the basis on cognitive explanation. From a logical point of view, such a concept can at best serve as a generic one, saying nothing about specific features of the phenomenon explained. In other words, for the theory of mind – philosophical or psychological – this concept, elaborated to the present extent, cannot be sufficient.

Further, I will try to elaborate on methodological shortcomings of classical representationalism, as I see them. In my view, the classical concept of representation comes from a primitive scheme of cognitive subject, surrounded by objects, which are mirrored in representations. But the whole concept of semantically laden extrinsic representations of objects is insufficiently substantiated and convincing; one could rather speak of sub-objective extrinsic representations: e.g., colour as the representation of a certain spectrum of reflected electro-magnetic radiation, etc. Akin to connectionism that introduces sub-symbol computations, it would be correct to talk about sub-symbol representations – extrinsic, as much as intrinsic: for instance, a vector of a neural network's weights may be seen as a representation of the categorical structure of data it has been trained on.

Then, representations have meaning only in the context of computations:

> *(Def:)* Structure *A* is a representation of structure *B* in the framework of a certain computation, if and only if, within this very framework, *A* and *B* are related by a stable invariant function.

By adopting such a 'weak' definition of representation, we find ourselves further away from the ultimate goal of cognitive theory, since so construed representations do not necessarily allow us to explain the process of obtaining knowledge in its complete form. But this is the only way to get rid of the 'homunculus' and see cognition as a process within a complex multi-level computing system.

Another complication may be linked to too anthropomorphic construal of 'semantic' relation of a representation to what is represented. As in the human world signs and their meanings are linked to each other conventionally, these links have to be *known* and, therefore, *taught* to humans to this end. If we borrow this kind of a semantic theory for a classical version of CS, we are at risk of colliding with another version of the homunculus paradox:

> *(HP2:)* For a symbolic computation to be semantically effective, the cognitive system must 'know' semantic relations between symbols and their references. But any knowledge is (based upon) representation. Thus, any representation needs another representation that supports it, and so on *ad infinitum*.

In the case of a purely syntactical computation, we avoid this paradox but leave the mechanism by which mental states generally have content (i.e., intentionality) unexplained. But if, having failed with computational accounts, we retreat to a purely and straightforwardly causal explications of the representing relation, we will eventually miss the point of the whole cognitive endeavour. So, this is another argument in favour of weakening the notion of representation for it to stay within CS as a useful explaining tool.

Representations in Neuroscience

Traditionally, psychology – cognitive as well as any other – is not directly determined by NS, sticking to functional or dynamical analyses and skipping mechanistic explanations. Therefore, any version of computational approach in psychology is a very promising move in connecting the two scientific realms. On the other hand, the digital-computational approach of the classical CS is of little help in this way, as it is not biologically realistic. Thus, what is needed there on the methodological level, is a shift from either dynamical approach or classical Turing-style computationalism to a generic computationalist theory that would comprehend the complicated architecture of neuronal computations. To this end, the cutting-edge cognitive (neuro)science is in need of a satisfactory mathematical theory applicable to natural, particularly neuronal, computations. Luckily, as Thompson and Piccinini put it, "<e>xperimental neuroscientists began talking about representations in the nervous system almost a century before the beginning of the cognitive revolution, which is so often associated with the contemporary dispute" [Thomson, Piccinini, 2018, p. 193].

With the latest advances in brain monitoring technologies, the ever growing amount of neuro-cognitive literature is now multiplied daily. But, as I have opted for a kind of multi-agent interpretation of knowledge, my special focus hereafter is put onto the research in the emerging field of cognitive social neuroscience. Nathan Emery [Emery, 2005] draws attention to the widespread opinion among researchers that life in a social group and predicting the behaviour of representatives of the same animal species require unprecedented levels of cognitive processing that are not displayed by non-primates. This 'social intelligence hypothesis' has been proposed as an alternative to more traditional explanations of the evolution of primates and human intelligence: such as using tools, hunting, extended spatial memory or mining industries.

In social cognitive neuroscience, the concept of the *Theory of Mind* (ToM) traditionally plays a significant role. ToM is also referred to as mentalisation, metarepresentation or secondary representation. It means the ability to understand the psychological or mental states of other people, such as their beliefs, desires and knowledge. The various forms of ToM are subdivided into three classes: the perceptual ToM (understanding of sight and attention), motivational (understanding of desires, goals and intentions) and informational (understanding of knowledge and beliefs) one. Thus, social cognition is interpreted as "the processing of any information which culminates in the accurate perception of the dispositions and intentions of other individuals" [Heberlein & Adolphs, 2005, p. 157].

The neurocognitive grounds of the self and self-awareness are explored in [Lieberman and Pfeifer, 2005]. Empirical data empower the argument that a special role in this regard is played by the posterior parietal cortex. This brain region is usually considered as important for the functions of maintaining working memory and spatial processing. However, this part of the brain can also be a place where non-symbolic, parallel, distributed representations are translated into symbolic, sequential, local representations.

I would like to highlight this circumstance specially. In my opinion, this is where the dividing line between individual and social cognitions is drawn. If the former depend entirely on the neural network architecture of the brain and are therefore parallel and distributed, the latter are formed in the course of social communication and depend (in the case of humans) on the architecture of the language, which is linear and consistent. And if the posterior parietal cortex is, indeed, the 'home' of self-consciousness, then this circumstance may well be considered as an empirical evidence for the 'self' being a social construct. Lieberman and Pfeifer point out that there is a temptation to think of the 'self' as an object with stable attributes. In fact, this temptation is not only for scientists, but for all people who value self-esteem and independence. However, as neurocognitive studies show, the 'self' is at least partially built and reconstructed over time as a function of situational and interpersonal constraints [Lieberman & Pfeifer, 2005, p. 223].

Conclusion

The review of principal issues related to conceptualisations of knowledge and its representation on three principal levels – conceptual (epistemology), modelling (cognitive science) and empirical (neuroscience) – bears the following results.

4.1. Out of the two competing epistemological approaches labelled as 'sole knower' and 'multi-agent', the latter is preferable, as it is not only better in explicating the modality of knowing, but, unlike its rival, provides a kind of mechanistic explanation of social knowledge processing.

4.2. The concept of representation initially provided by cognitive science at its classical stage is no longer satisfactory in view of the latest developments in the sciences based on the computational approaches. Representations must be re-construed as relative and instant aspects of computations of various kinds taking place in the nature, in the mind and in the society.

4.3. Empirical evidence provided by social cognitive neuroscience reveal functional brain regions directly engaged in the social intercourse of its owners. The important fact is that one and the same region is responsible for both self-consciousness and the symbolic activity. This makes for a plausible hypothesis that self-consciousness, 'theory of mind' and linguistic capacities are virtually the cognitive foundations of social being.

The general conclusion is that the cutting-edge science of mind and knowledge, as well as all the sciences of the human being, is in need of an updated theory of computation that would embrace neural, cognitive and social realms altogether.

Список литературы / References

Anderson, 1983 – Anderson, J.R. *The Architecture of Cognition.* Cambridge, Massachusetts: Harvard University Press, 1983, 340 pp.

Attneave, 1961 – Attneave, F. "In Defence of Homunculi", in: Rosenblith, W.A. (ed.) *Sensory communication: Contributions to the symposium on principles of sensory communication, July 19 – Aug. 1*, 1959, Endicott House, M. I. T. 1961, pp. 777–782

Emery, 2005 – Emery, N.J. The Evolution of Social Cognition", in: A. Easton and N. J. Emery (eds). *The Cognitive Neuroscience of Social Behavior.* New York: Psychology Press, 2005, pp. 115–156.

Fagin et al., 1995 – Fagin, R; Halpern, J.Y.; Moses, Y. & Vardi, M.Y. *Reasoning About Knowledge.* The MIT Press: Cambridge, Massachusetts; London, England, 1995, 515 pp.

Fodor, 1975 – Fodor, J.A. *The Language of Thought.* New York: Thomas Y. Crowell Company, 1975, 214 pp.

Fodor, 2008 – Fodor J. LOT 2: The Language of Thought Revisited. New York: Oxford University Press. 2008. 228 pp.

Fresco 2012 – Fresco, N. "The Explanatory Role of Computation in Cognitive Science", *Minds & Machines*, 2012, no. 22, pp. 353–380.

Heberlein & Adolphs, 2005 – Heberlein, A.S. & Adolphs, R. "Functional Anatomy of Human Social Cognition", in: A. Easton & N.J. Emery (eds). *The Cognitive Neuroscience of Social Behavior*., New York: Psychology Press, 2005, pp. 157–194.

Henderson & Horgan, 2011 – Henderson, D. & Horgan, T. *The Epistemological Spectrum. At the Interface of Cognitive Science and Conceptual Analysis.* Oxford University Press: Oxford, 2011, 292 pp.

Hutto & Myin, 2013 – Hutto, D.D. & Myin, E. *Radicalizing Enactivism: Basic Minds Without Content.* Cambridge: MIT Press, 2013, 206 pp.

Hutto, 2011 – Hutto, D. D. "Representation Reconsidered. Representation Reconsidered", *Philosophical Psychology*, 2011, vol. 24, no. 1, pp. 135–139.

Kovalyov & Rodin, 2018 – Kovalyov, S.P. & Rodin, A.V. "Problema obosnovanija v formal'nom predstavlenii znaniy" [The Problem of Justification in Formal Knowledge Representation], *Vestnik Tomskogo gosudarstvennogo universiteta. Filosofiya. Sotsiologiya. Politologiya – Tomsk State University Journal of Philosophy, Sociology and Political Science*, 2018, vol. 46, pp. 14–29. (In Russian)

Lieberman & Pfeifer, 2005 – Lieberman, M.D. & Pfeifer, J.H. "The Self and Social Perception: Three Kinds of Questions in Social Cognitive Neuroscience", in: A. Easton & N.J. Emery (eds). *The Cognitive Neuroscience of Social Behavior*., New York: Psychology Press, 2005, pp. 195–235.

MacLennan, 2004 – MacLennan, B.J. "Natural Computation And Non-Turing Models Of Computation", *Theoretical Computer Science*, 2004, vol. 317, pp. 115–145.

Marr, 2010 – Marr, D. *Vision: a Computational Investigation into the Human Representation and Processing of Visual Information.* Cambridge: MIT Press, 2010, 369 pp.

Nikiforov, 2018 – Nikiforov, A.L. “Chto takoe znanie? Poiski opredeleniya” [What is Knowledge? A Search for Definition], in: I.T. Kasavin & N.N. Voronina (eds.). Epistemologiya segodnya. Idei, problemy, diskussii: monografija [Epistemology Today. Ideas, Problems, Discussions: The Monography]. Nizhnii Novgorod: NN State University Press, 2018, 413 pp. (In Russian)

Piccinini, 2018 – Piccinini, G. “Computation and Representation in Cognitive Neuroscience”, *Minds & Machines*, 2018, vol. 28, iss. 1, pp. 1–6.

Poggio, 2012 – Poggio, T. “The Levels of Understanding Framework, Revised”, *Perception*, 2012, vol. 41, pp. 1017–1023.

Thomson & Piccinini, 2018 – Thomson, E. & Piccinini, G. “Neural Representations Observed”, *Minds and Machines*, 2018, vol. 28(1), pp. 191–235.

Veber, 2014 – Veber, M. “Knowledge With and Without Belief”, *Metaphilosophy*, 2015, vol. 45, no. 1, pp. 120–132.

Vlassis, 2007 — Vlassis, N.A. *Concise Introduction to Multiagent Systems and Distributed Artificial Intelligence.* New York: Morgan & Claypool, 2007, 71 pp.

Wooldridge, 2002 – Wooldridge, M. *An Introduction to Multiagent Systems.* Chichester: JohnWiley & Sons Ltd., 2002, 338 pp.